Département Archivage & Diffusion CINES

[VITAM R7 - PAC V3]

Rapport d'étude fonctionnelle comparative

Table des matières

Introduction 3
Méthodologie 4
Synthèse de l'analyse 8
Bilan des résultats par entité fonctionnelle 8
Success story Le POC CROUS 10
Définition des concepts fonctionnels VITAM 11
Etude détaillée par axes et entités fonctionnelles associées 13
Entrée 13
Gestion des archives existantes 18
Accès 19
Préservation 20
Administration fonctionnelle 21
Administration technique 21
Stockage et Gestion des données 22
Transverse 23
[VITAM-PAC] Récapitulatif 25

Introduction

VITAM (Valeurs Immatérielles Transmises aux Archives pour Mémoire) est un programme interministériel visant à développer un socle d'archivage électronique réutilisable par l'ensemble des administrations françaises pour classer, conserver et sécuriser les documents numériques.

Fruit de la collaboration de trois ministères (Affaires Etrangères et Développement international, Culture et Communication, Défense), ce chantier informatique est piloté par la direction interministérielle des systèmes d'information et de communication (DINSIC), au sein du secrétariat général pour la modernisation de l'action publique (SGMAP).

Le CINES a une mission stratégique au niveau de l'archivage pérenne des données électroniques du monde de l'Enseignement Supérieur et de la Recherche. Dans ce cadre, il a obtenu l'agrément du Service interministériel des Archives de France (SIAF) qui lui permet de proposer des services de conservation d'archives numériques publiques intermédiaires.

Le CINES possède une plateforme d'archivage électronique – PAC – qui est exploitée depuis plus de 10 ans. Elle est constituée d'un logiciel propriétaire ARCSYS (société Infotel) avec une surcouche développée en interne permettant d'assurer tous les traitements de pérennisation à l'entrée des SIP.

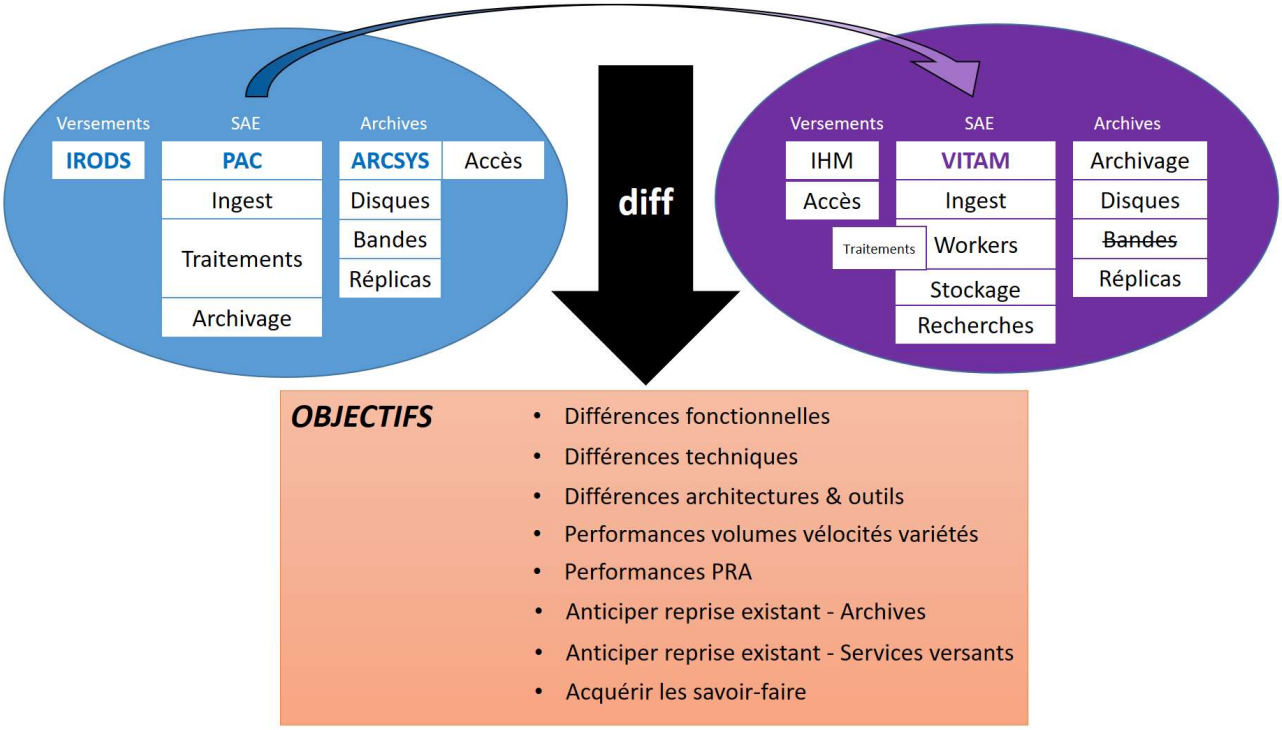
Fort d'une riche expérience en matière de pérennisation des informations numériques et d'un intérêt pour une éventuelle réutilisation des briques logicielles produites par ce programme interministériel, le CINES a signé le 30 juin 2016 une convention de partenariat avec la DINSIC et VITAM.

Dans un but de veille technologique, il a semblé pertinent au CINES d'évaluer la couverture fonctionnelle de VITAM par rapport à celle de PAC.

En parallèle de cette étude comparative, un démonstrateur en situation réelle de production pour le Crous de Montpellier en tant que service versant a été réalisé sur cette même période.

Méthodologie

L'objectif du CINES est d'étudier le spectre fonctionnel et technique du SAE VITAM en comparaison du SAE actuel. Les différents axes de cette étude sont explicités ci-dessous :

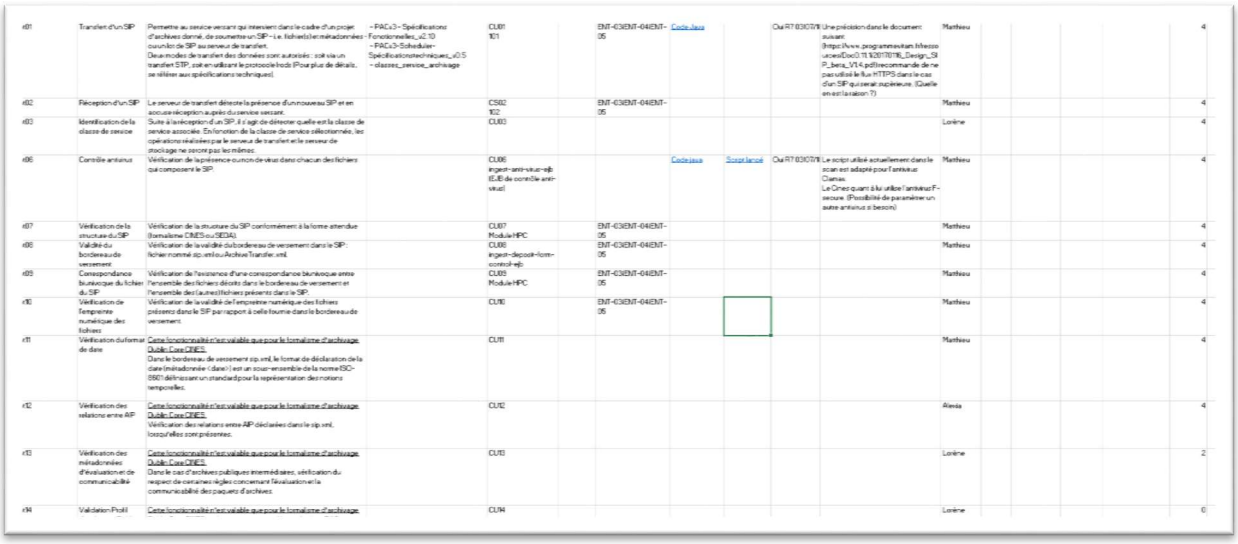


Elle a été réalisée dans les conditions suivantes :

Durée	six mois
Acteurs	une équipe de treize personnes de profils variés (Archivistes / Ingénieurs en développement / Architectes / Ingénieurs réseau et système d'information …)
Versions concernées	la R7 de Vitam, livrée en milieu d'année 2018
Méthode	le Kanban, une des méthodes Agile

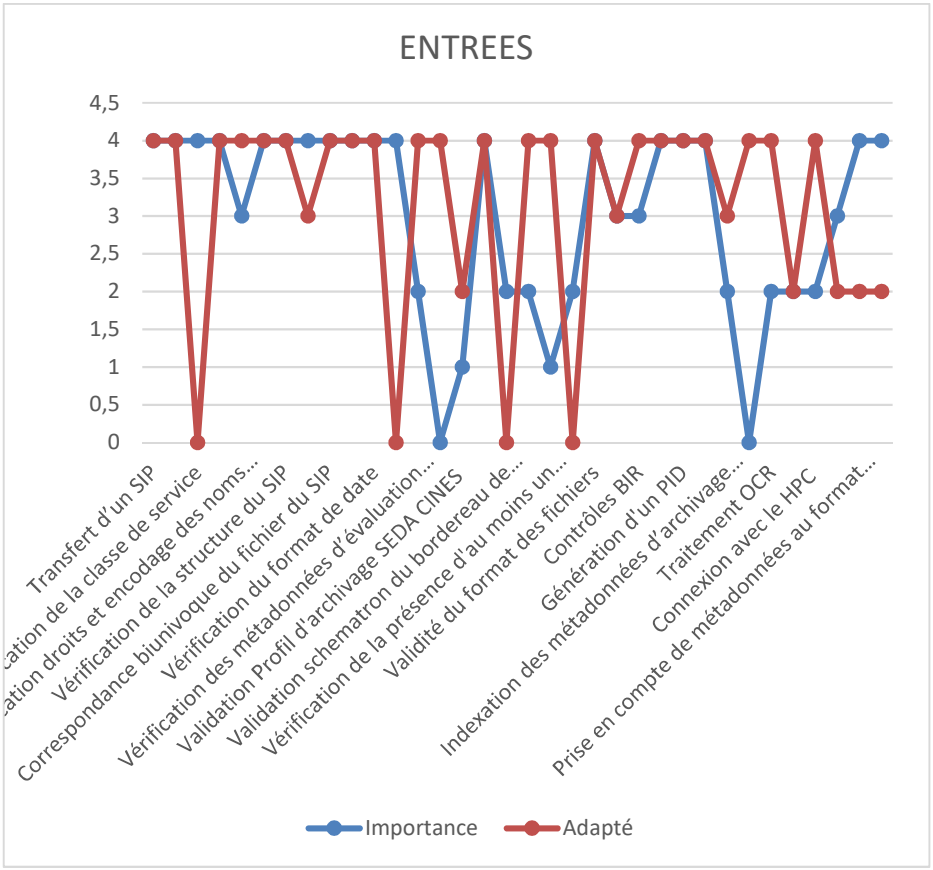
La structure de ce rapport, inspirée du modèle conceptuel OAIS, est composée d'une comparaison de sept groupes de fonctionnalités entre les deux SAE, Vitam d'une part et la plateforme PAC (Plateforme d'Archivage du CINES) d'autre part.

Ci-dessous, un extrait du « Backlog » des exigences du CINES qui a été réalisé dans le cadre de la méthode Kanban.

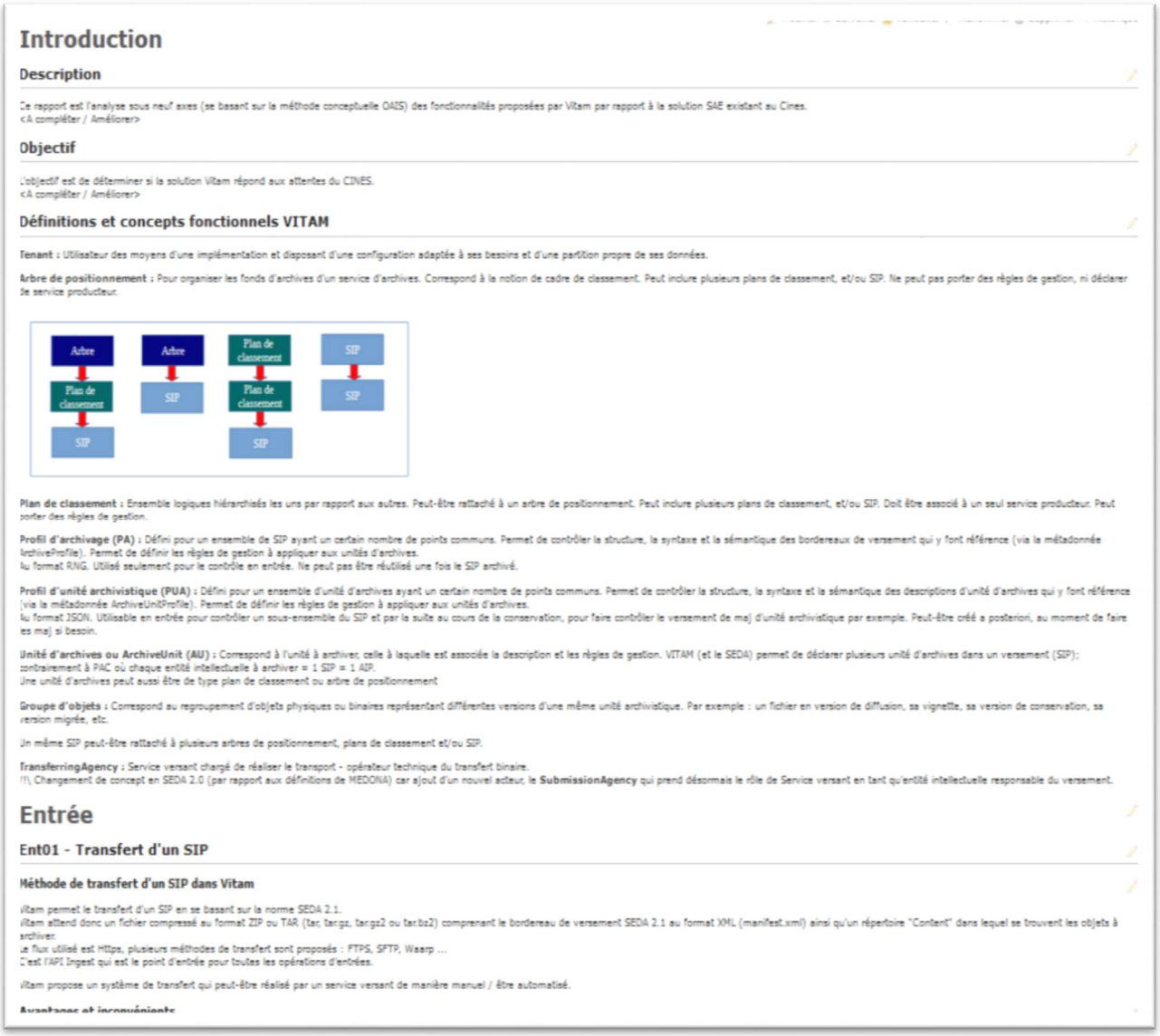


r01	Transfert d'un SIP	Permettre au service versant qui intervient dans le cadre d'un projet d'archivage donné, de soumettre un SIP – i.e. fichier(s) et métadonnées – ou un lot de SIP au serveur de transfert. Deux modes de transfert des données sont autorisés : soit via un transfert SFTP, soit en utilisant le protocole Http (Pour plus de détails, se référer aux spécifications techniques).	- PAC1-Spécifications Fonctionnelles_v2.10 - PAC1-Scheduler-Spécificationstechniques_v0.5 - classes_service_archivage	CU01 101	ENT-03ENT-04ENT-05	Code Java		Oui RT 03078	Une précision dans le document suivant (https://www.programmevitam.fr/ressources/Doc/0.11.1/20170116_Design_SIP_beta_V1.4.pdf) recommande de ne pas utiliser le flux HTTPS dans le cas d'un SIP qui serait supérieure. (Quelle en est la raison ?)	Matthieu		4
r02	Réception d'un SIP	Le serveur de transfert détecte la présence d'un nouveau SIP et en accuse réception auprès du service versant.		CS02 102	ENT-03ENT-04ENT-05					Matthieu		4
r03	Identification de la classe de service	Suite à la réception d'un SIP, il s'agit de déterminer quelle est la classe de service associée. En fonction de la classe de service sélectionnée, les opérations réalisées par le serveur de transfert et le serveur de stockage ne seront pas les mêmes.		CU03						Lorène		4
r06	Contrôle antivirus	Vérification de la présence ou non de virus dans chacun des fichiers qui composent le SIP.		CU06 ingest-anti-virus-ejb (EJB de contrôle anti-virus)		Code Java	Script lancé	Oui RT 03078	Le script utilisé actuellement dans le scan est adapté pour l'antivirus Clamav. Le Cines quant à lui utilise l'antivirus F-secure. (Possibilité de paramétrer un autre antivirus si besoin)	Matthieu		4
r07	Vérification de la structure du SIP	Vérification de la structure du SIP conformément à la forme attendue (formalisme CINES ou SEDA).		CU07 Module HPC	ENT-03ENT-04ENT-05					Matthieu		4
r08	Validité du bordereau de versement	Vérification de la validité du bordereau de versement dans le SIP: fichier nommé sip.xml ou Archive Transfer.xml.		CU08 ingest-depot-form-control-ejb	ENT-03ENT-04ENT-05					Matthieu		4
r09	Correspondance biunivoque du fichier du SIP	Vérification de l'existence d'une correspondance biunivoque entre l'ensemble des fichiers décrits dans le bordereau de versement et l'ensemble des (autres) fichiers présents dans le SIP.		CU09 Module HPC	ENT-03ENT-04ENT-05					Matthieu		4
r10	Vérification de l'empreinte numérique des fichiers	Vérification de la validité de l'empreinte numérique des fichiers présents dans le SIP par rapport à celle fournie dans le bordereau de versement.		CU10	ENT-03ENT-04ENT-05					Matthieu		4
r11	Vérification du format de date	Cette fonctionnalité n'est valable que pour le formalisme d'archivage Dublin Core CINES. Dans le bordereau de versement sip.xml, le format de déclaration de la date (métadonnée <date>) est un sous-ensemble de la norme ISO-8601 définissant un standard pour la représentation des notions temporelles.		CU11						Matthieu		4
r12	Vérification des relations entre AIP	Cette fonctionnalité n'est valable que pour le formalisme d'archivage Dublin Core CINES. Vérification des relations entre AIP déclarées dans le sip.xml, lorsqu'elles sont présentes.		CU12						Alexis		4
r13	Vérification des métadonnées d'évaluation et de communicabilité	Cette fonctionnalité n'est valable que pour le formalisme d'archivage Dublin Core CINES. Dans le cas d'archives publiques intermédiaires, vérification du respect de certaines règles concernant l'évaluation et la communicabilité des paquets d'archives.		CU13						Lorène		2
r14	Validation Profil	Cette fonctionnalité n'est valable que pour le formalisme d'archivage		CU14						Lorène		0

Les fonctionnalités présentes dans le SAE du CINES ont été notées par importance sur une échelle de 0 à 4 (0 - Nulle, 1 – Mineure, 2 – Modérée, 3 – Majeure, 4- Obligatoire) puis comparées avec celles proposées par Vitam pour savoir si elles étaient adaptées (0 – Non, 1 – Peu, 2 – Moyen, 3 – Bien, 4 – Adapté), l'objectif étant de déterminer si la solution Vitam pourrait répondre aux attentes du CINES sur chaque axe.



Un premier rapport a été rédigé sur un wiki pour détailler, exigence par exigence, le résultat de l'analyse. En voici un extrait :



Introduction

Description

Ce rapport est l'analyse sous neuf axes (se basant sur la méthode conceptuelle OAIS) des fonctionnalités proposées par Vitam par rapport à la solution SAE existant au Cines.
<A compléter / Améliorer>

Objectif

L'objectif est de déterminer si la solution Vitam répond aux attentes du CINES.
<A compléter / Améliorer>

Définitions et concepts fonctionnels VITAM

Tenant : Utilisateur des moyens d'une implémentation et disposant d'une configuration adaptée à ses besoins et d'une partition propre de ses données.

Arbre de positionnement : Pour organiser les fonds d'archives d'un service d'archives. Correspond à la notion de cadre de classement. Peut inclure plusieurs plans de classement, et/ou SIP. Ne peut pas porter des règles de gestion, ni déclarer de service producteur.

Plan de classement : Ensemble logiques hiérarchisés les uns par rapport aux autres. Peut-être rattaché à un arbre de positionnement. Peut inclure plusieurs plans de classement, et/ou SIP. Doit être associé à un seul service producteur. Peut porter des règles de gestion.

Profil d'archivage (PA) : Défini pour un ensemble de SIP ayant un certain nombre de points communs. Permet de contrôler la structure, la syntaxe et la sémantique des bordereaux de versement qui y font référence (via la métadonnée ArchiveProfile). Permet de définir les règles de gestion à appliquer aux unités d'archives.
Au format RNG. Utilisé seulement pour le contrôle en entrée. Ne peut pas être réutilisé une fois le SIP archivé.

Profil d'unité archivistique (PUA) : Défini pour un ensemble d'unité d'archives ayant un certain nombre de points communs. Permet de contrôler la structure, la syntaxe et la sémantique des descriptions d'unité d'archives qui y font référence (via la métadonnée ArchiveUnitProfile). Permet de définir les règles de gestion à appliquer aux unités d'archives.
Au format JSON. Utilisable en entrée pour contrôler un sous-ensemble du SIP et par la suite au cours de la conservation, pour faire contrôler le versement de maj d'unité archivistique par exemple. Peut-être créé a posteriori, au moment de faire les maj si besoin.

Unité d'archives ou ArchiveUnit (AU) : Correspond à l'unité à archiver, celle à laquelle est associée la description et les règles de gestion. VITAM (et le SEDA) permet de déclarer plusieurs unité d'archives dans un versement (SIP); contrairement à PAC où chaque entité intellectuelle à archiver = 1 SIP = 1 AIP.
Une unité d'archives peut aussi être de type plan de classement ou arbre de positionnement

Groupe d'objets : Correspond au regroupement d'objets physiques ou binaires représentant différentes versions d'une même unité archivistique. Par exemple : un fichier en version de diffusion, sa vignette, sa version de conservation, sa version migrée, etc.

Un même SIP peut-être rattaché à plusieurs arbres de positionnement, plans de classement et/ou SIP.

TransferringAgency : Service versant chargé de réaliser le transport - opérateur technique du transfert binaire.
(!) Changement de concept en SEDA 2.0 (par rapport aux définitions de MEDONA) car ajout d'un nouvel acteur, le **SubmissionAgency** qui prend désormais le rôle de Service versant en tant qu'entité intellectuelle responsable du versement.

Entrée

Ent01 - Transfert d'un SIP

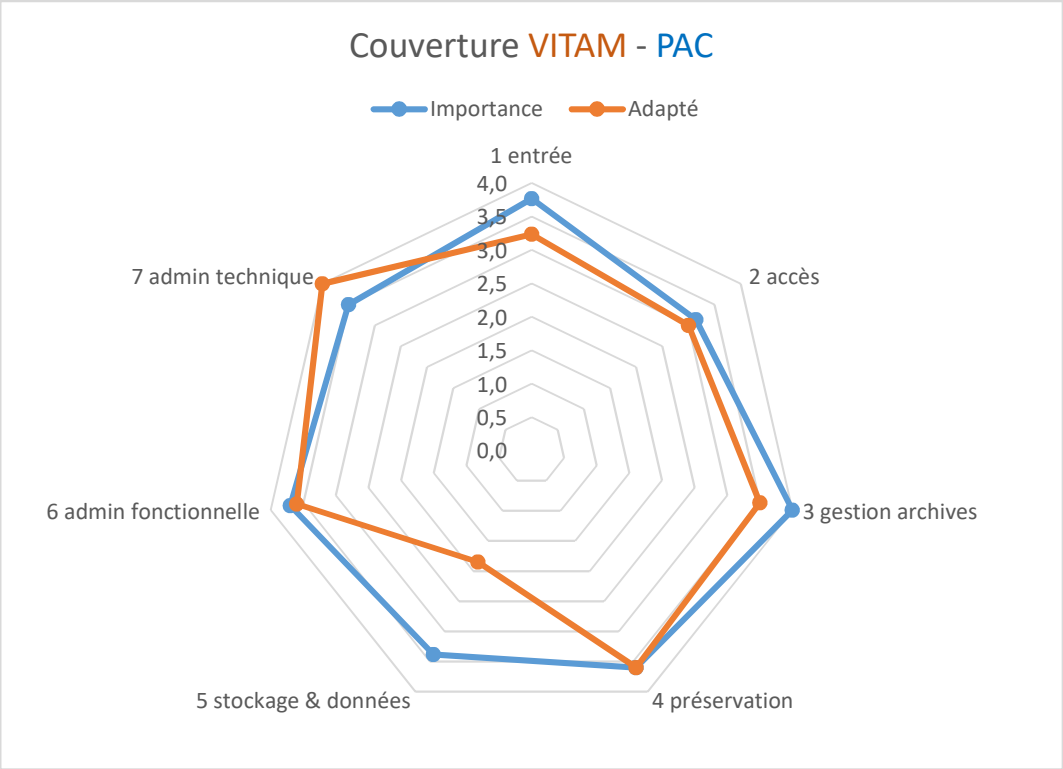
Méthode de transfert d'un SIP dans Vitam

Vitam permet le transfert d'un SIP en se basant sur la norme SEDA 2.1.
Vitam attend donc un fichier compressé au format ZIP ou TAR (tar, tar.gz, tar.gz2 ou tar.bz2) comprenant le bordereau de versement SEDA 2.1 au format XML (manifest.xml) ainsi qu'un répertoire "Content" dans lequel se trouvent les objets à archiver.
Le flux utilisé est Https, plusieurs méthodes de transfert sont proposés : FTPS, SFTP, Waarp ...
C'est l'API Ingest qui est le point d'entrée pour toutes les opérations d'entrées.

Vitam propose un système de transfert qui peut-être réalisé par un service versant de manière manuel / être automatisé.

Avantages et inconvénients

Le résultat final de cette étude peut se résumer grâce au graphe ci-dessous :



	AXES	Importance	Adapté	%
1	entrée	3,0	3,1	79%
2	accès	3,1	3,0	75%
3	gestion archives	4,0	3,5	88%
4	préservation	3,6	3,6	90%
5	stockage & données	3,4	1,8	46%
6	admin fonctionnelle	3,7	3,6	90%
7	admin technique	3,5	4,0	100%
8	RGPD	4,0	4,0	100%
9	Architecture	4,0	4,0	100%
	Total	**3,6**	**3,4**	**85,3%**

Synthèse de l'analyse

Le premier résultat de cette étude est positif : l'outil répond correctement aux besoins archivistiques, car il a été conçu et élaboré par des archivistes pour des archivistes.

Certaines fonctionnalités, indispensables pour une utilisation potentielle de l'outil au CINES, restent encore à développer ou sont en cours de développement, ce qui peut expliquer sur certains axes un écart entre l'importance du critère pour le CINES et l'adaptation de VITAM. Un résumé des résultats par axe est présenté ci-après. L'analyse plus complète de l'ensemble des fonctionnalités testées est détaillée dans une seconde partie de ce rapport.

Les travaux menés ont également envisagé la manière de reprendre l'arriéré des données déjà archivées au CINES afin de les migrer dans l'instance VITAM, qui serait déployée au CINES. Ces réflexions doivent encore être poursuivies au regard des fonctionnalités attendues.

La règlementation relative à la protection des données personnelles (RGDP) est prise en compte dans le périmètre du back office de VITAM de manière à permettre sa mise en œuvre dans les institutions.

Bilan des résultats par entité fonctionnelle

Entrées :

La fonctionnalité d'Entrées constitue un des éléments prépondérants dans le processus d'archivage numérique mis en place au CINES (38% des fonctionnalités analysées).

Le besoin est couvert dans sa majorité. Toutefois, certains points restent à surveiller, car toutes les fonctionnalités ne sont pas encore livrées, et plusieurs évolutions sont à développer pour s'ajuster au contexte du CINES.

Le formalisme d'archivage utilisé par VITAM est limité à la dernière version du SEDA (v2.1). Or, le CINES utilise d'autres formalismes d'archivage (SEDA v1.0 et Dublin Core) qu'il souhaite conserver.

Un outil est nécessaire pour convertir, en amont de l'entrée, l'ensemble des SIP pour les rendre compatibles avec VITAM. Cela pourra notamment occasionner un travail conséquent de rédaction de profils d'archivage et de profils d'unité archivistique, pour lesquels les outils d'aide à la création n'existent pas encore ou dans des versions difficiles à utiliser.

Le CINES effectue plusieurs traitements spécifiques (externes à VITAM) lors du processus d'ingestion des SIP.

Pour ces traitements ainsi que pour les contrôles d'entrées, le besoin est de pouvoir les paramétrer en fonction de classes de service préalablement définies.

Les fonctionnalités permettant la couverture de ces besoins sont en cours de livraison côté VITAM et devront être testées finement pour valider leur utilisation.

Le CINES doit enfin repenser ses pratiques actuelles afin d'exploiter au mieux les possibilités offertes par l'application Vitam.

Les réflexions à avoir portent majoritairement sur la granularité entre les SIP et les AIP, la manière d'exprimer les relations entre les unités d'archives ou encore de lier les SIP à leur convention d'archivage.

Accès

La fonctionnalité d'Accès répond aux attentes du CINES en offrant de larges possibilités de paramétrage. Il serait nécessaire, pour la suite, de définir les besoins précis pour communiquer au mieux avec le ou les front d'accès.

Gestion des archives

La fonctionnalité de Gestion des archives couvre correctement les besoins du CINES.

Il s'agit de fonctionnalités majoritairement archivistiques (administration des règles de gestion et des éliminations notamment) qui sont bien traitées dans Vitam.

La fonctionnalité de Transfert vers un autre SAE étant en cours de spécifications par les équipes VITAM, il convient de suivre au plus près son développement.

Préservation

La fonctionnalité de Préservation a été récemment implémentée dans VITAM suite aux besoins exprimés par le CINES.

Des tests et paramétrages devront être réalisés afin de déterminer si les développements livrés, de cette fonctionnalité, correspondent pleinement aux besoins du CINES.

Stockage et gestion des données

Peu de tests ont été réalisés lors de l'étude, car la version testée ne proposait que peu de possibilités sur les fonctionnalités liées au stockage de données.

Il est prévu d'expérimenter l'implémentation du stockage sur bandes en partenariat avec les équipes de Vitam.

D'autres tests sur l'évaluation d'offre de stockage objet seront à prévoir pour étudier la faisabilité et la pertinence d'utilisation de ce type d'offre au CINES.

Administrations fonctionnelle et technique

Les fonctionnalités d'Administrations fonctionnelle et technique sont très larges et répondent correctement aux besoins du CINES.

Un travail important devra être effectué pour choisir la configuration la plus appropriée. Des scenarii sont actuellement en cours de test sur le sujet, ils permettront de déterminer les besoins techniques et fonctionnels nécessaires au développement d'un front d'administration du back office VITAM.

Des axes d'amélioration sont à prévoir sur les aspects d'intégration et d'exploitation de la solution. Le CINES est en discussion avec les équipes VITAM sur ces points.

Success story

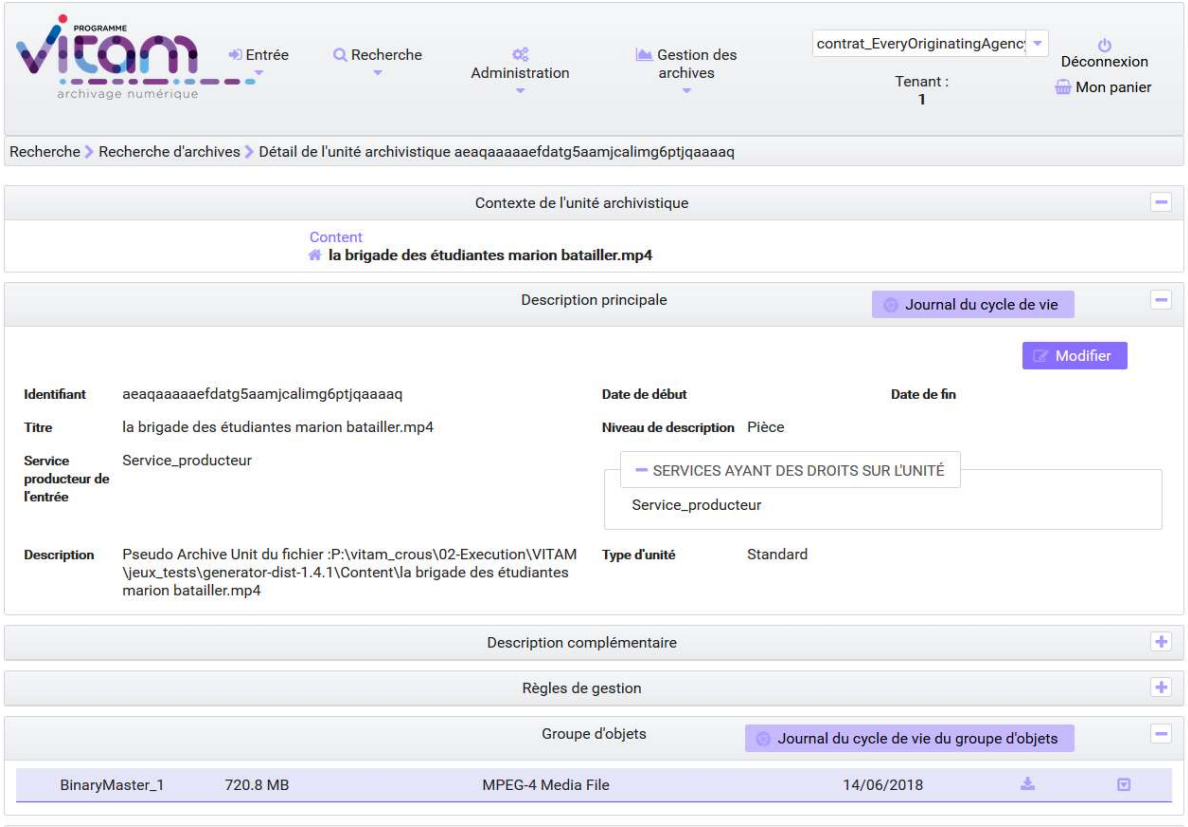
Le POC CROUS

Le périmètre de ce démonstrateur de faisabilité est l'archivage des données defournies par le service culturel de CORUM (Ged Nuxeo du Crous de Montpellier). Il s'agit, plus précisément, des œuvres culturelles qui ont remporté les différents concours étudiants (Film, Essai, BD, Photo, Créations, etc.).

Le CROUS de Montpellier a réalisé avec l'aide de la société ATOS et sous la gestion de projet du CINES, un connecteur pour sa GED ainsi qu'une interface web de suivi des versements.

L'objectif du POC était de réussir à automatiser le versement des données culturelles du CROUS dans le SAE expérimental VITAM (version « médium » avec 15 machines virtuelles), déployé dans l'environnement de production du CINES.

Les versements et l'accès aux données vers et depuis le SAE ont été réalisés avec succès, prouvant ainsi la capacité de VITAM à fonctionner en production avec un vrai service versant.



Définition des concepts fonctionnels VITAM

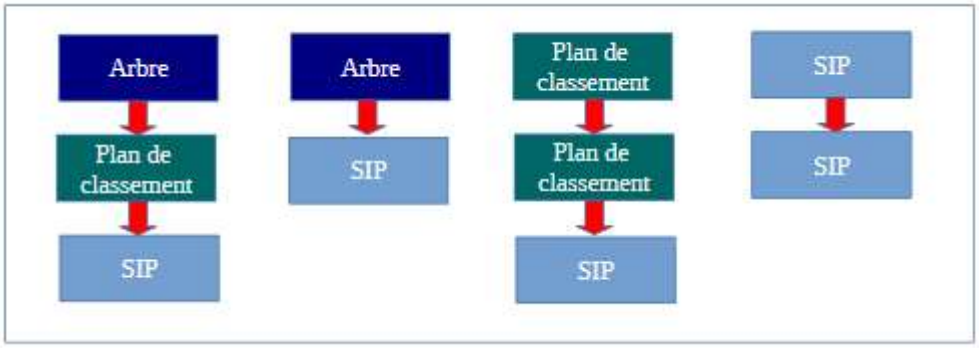
Tenant : Utilisateur des moyens d'une implémentation et disposant d'une configuration adaptée à ses besoins et d'une partition propre de ses données.

SIP (Submission Information Package) : correspond au paquet d'archives à verser.

Arbre de positionnement : Pour organiser les fonds d'archives d'un service d'archives. Correspond à la notion de cadre de classement.

Peut inclure plusieurs plans de classement, et/ou SIP.

Ne peut pas porter de règles de gestion, ni déclarer de service producteur.



Plan de classement : *Ensemble logiques hiérarchisés les uns par rapport aux autres*. Peut-être rattaché à un arbre de positionnement.

Peut inclure plusieurs plans de classement, et/ou SIP.

Doit être associé à un seul service producteur.

Peut porter des règles de gestion.

Profil d'archivage (PA) :

- Défini pour un ensemble de SIP ayant un certain nombre de points communs.
- Permet de contrôler la structure, la syntaxe et la sémantique des bordereaux de versement qui y font référence (via la métadonnée ArchiveProfile).
- Permet de définir les règles de gestion à appliquer aux unités d'archives.
- Au format RNG. Utilisé seulement pour le contrôle en entrée.
- Ne peut pas être réutilisé une fois le SIP archivé.

Profil d'unité archivistique (PUA) :

- Défini pour un ensemble d'unité d'archives ayant un certain nombre de points communs.
- Permet de contrôler la structure, la syntaxe et la sémantique des descriptions d'unité d'archives qui y font référence (via la métadonnée ArchiveUnitProfile).
- Permet de définir les règles de gestion à appliquer aux unités d'archives.
- Au format JSON. Utilisé en entrée pour contrôler un sous-ensemble du SIP et par la suite au cours de la conservation, pour faire contrôler le versement de mise à jour d'unité archivistique par exemple.
- Peut-être créé a posteriori, au moment de faire les mises à jour si besoin.

Unité d'archives ou ArchiveUnit (AU) :

- Correspond à l'unité à archiver, celle à laquelle est associée la description et les règles de gestion. VITAM (et le SEDA) permettent de déclarer plusieurs unité d'archives dans un versement (SIP); contrairement à PAC où chaque entité intellectuelle à archiver = 1 SIP = 1 AIP.
- Une unité d'archives peut aussi être de type plan de classement ou arbre de positionnement.

Groupe d'objets :

- Ensemble d'objets physiques ou binaires représentant différentes versions d'une même unité archivistique. Par exemple : un fichier en version de diffusion, sa vignette, sa version de conservation, sa version migrée, etc.
- Un même SIP peut-être rattaché à plusieurs arbres de positionnement, plans de classement et/ou SIP.

TransferringAgency :

- Service versant chargé de réaliser le transport - opérateur technique du transfert binaire.
- /!\ Changement de concept en SEDA 2.0 (par rapport aux définitions de MEDONA) car ajout d'un nouvel acteur, le **SubmissionAgency** qui prend désormais le rôle de Service versant en tant qu'entité intellectuelle responsable du versement.

Etude détaillée par axes et entités fonctionnelles associées

Entrée

Ent01 - Transfert d'un SIP

Méthode de transfert d'un SIP dans Vitam

- Vitam permet le transfert d'un SIP en se basant sur la norme SEDA 2.1
- Vitam attend donc un fichier compressé au format ZIP ou TAR (tar, tar.gz, tar.gz2 ou tar.bz2) comprenant le bordereau de versement SEDA 2.1 au format XML (manifest.xml) ainsi qu'un répertoire "Content" dans lequel se trouvent les objets à archiver.
- Le flux utilisé est Https, plusieurs méthodes de transfert sont proposées : FTPS, SFTP, Waarp ...
- C'est l'API Ingest qui est le point d'entrée pour toutes les opérations d'entrées.
- Vitam propose un système de transfert qui peut être réalisé par un service versant de manière manuelle / automatisée.

Avantages et inconvénients

+ : Possibilité de connecter Irods à Vitam.

Ent02 - Réception d'un SIP

Réception d'un SIP dans VITAM

- La réception d'un SIP se fait dans Vitam dès son envoi par le service versant.
- Une notification en http est envoyée automatiquement au service versant pour préciser le statut de l'import du SIP.
- Trois types de retours possibles : OK, KO, FATAL

Divergence :

- On note, à l'inverse de PAC, qu'aucune notification mail n'est réalisée par Vitam.

Ent03 - Identification de la classe de service

- Cette fonctionnalité est liée à la possibilité de paramétrer le workflow d'entrée en sélectionnant les contrôles à faire manuellement ou via des plugins externes.
- Il n'est pas possible, actuellement, de réaliser un niveau de classe de service avec Vitam (workflow en dur dans le code Vitam et non personnalisable).
- Le besoin de proposer plusieurs niveaux de service a été remonté aux équipes Vitam : possible et en cours d'étude (certainement en lien avec le paramétrage des contrats d'entrée).

Evolutions réalisées suite aux ateliers avec Vitam en Décembre 2018 :

- Une modification de Vitam a été réalisée dans la V2, livrée le 18/02, concernant la possibilité d'utiliser des « griffons ».
- Ces griffons sont utilisés dans le cadre de la préservation des données mais il est également prévu par Vitam, dans leur roadmap de pouvoir les activer à l'entrée.

- Certains tests devraient être réalisés sur l'utilisation de plugin/griffon.

Ent04 – Unicité du versement

- Contrairement à PAC qui réalise, dès l'entrée, un contrôle d'unicité des archives et envoie une alerte de doublon si le SIP existe déjà dans le référentiel des AIP, Vitam ne prévoit pas ce type de contrôle.

Suite aux ateliers avec Vitam en Décembre 2018 :

- Il est prévu dans la feuille de route de Vitam d'étudier le traitement des doublons de versement d'archive.

Ent05 – Vérification des droits et encodage des noms de fichiers

- Vitam réalise des contrôles sur l'encodage de nom de fichier qui ne doit comprendre ni caractère accentué, ni virgule, ni apostrophe, ni parenthèse, ni espace.
- Le nommage du fichier doit respecter la REGEX suivante : [a-zA-Z0-9.\\-\\/+=@_]*$.
- La longueur de nom de fichier (répertoire + nom) est également limitée à 256 caractères.
- Les contrôles réalisés par VITAM sur les droits et l'encodage des noms des fichiers sont similaires à ceux réalisés par PAC.

Ent06 - Contrôle Antivirus

- Vitam réalise un contrôle sur l'intégralité des fichiers qui composent le SIP, tandis que PAC arrête le scan de l'antivirus dès détection d'un fichier infecté.
- Vitam propose Clamav (open source) comme antivirus par défaut, tandis que le CINESutilise l'antivirus F-secure qui nécessite une licence.A ce jour, les fichiers qui dépassent les 2 Go ne peuvent pas être analysés par le scan de l'antivirus F-secure (car logiciel 32 bit), tandis que la solution Clamav n'a pas de limitation de taille pour le scan des fichiers.
- Il est important de préciser que Vitam peut être configuré avec n'importe quel antivirus.

Ent07 : Vérification de la structure du SIP

- Les contrôles réalisés par VITAM sur la structure du SIP sont similaires à ceux réalisés par PAC.

Divergences

- Le CINES qui utilise actuellement un formalisme de SIP spécifique dont le bordereau de versement repose en partie sur du Dublin Core et le formalisme du SEDA 1.0, n'envisage pas d'abandonner ces deux formalismes en entrée pour s'adapter à Vitam qui n'accepte que des paquets conformes à la norme SEDA 2.1.
- Le CINES prévoit le développement d'un outil de conversion de SIP conformes à VITAM (SEDA 2.1) à partir de SIP CINES en DublinCore et SEDA 1.0. Cet outil interviendrait avant l'entrée dans Vitam.

Ent08 - Validité du bordereau de versement

- Les contrôles réalisés par VITAM concernant le bordereau de versement et les comportements en cas de rejet sont similaires à ceux réalisés par PAC.

Ent09 - Correspondance biunivoque du fichier du SIP

- Les contrôles réalisés par VITAM sur ce point sont similaires à ceux réalisés par PAC.

Ent10 - Vérification de l'empreinte numérique des fichiers

- Les contrôles réalisés par VITAM sur ce point sont similaires à ceux réalisés par PAC.

Ent11 - Vérification du format de date

- Pour le formalisme d'archivage spécifique au CINES basé sur du Dublin Core, les contrôles effectués par PAC sur la métadonnée « Date » sont plus poussés que ceux de VITAM : possibilité d'avoir AAAA-MM, AAAA seulement, le siècle (AA) ou la valeur « s.d. ».
- Le SEDA 2.1 est plus restrictif au niveau des formes acceptées : seulement AAAA-MM-JJ, pourtant la norme est bien la même (ISO 8601).
- Une solution de contournement a été trouvée, testée et validée, à savoir utiliser le profil d'unité archivistique et l'ontologie pour créer une autre métadonnée dans le SEDA, avec un type string et un pattern associés pour lister les valeurs possibles.

Ent12 - Vérification des relations entre AIP

- L'équipe de Vitam confirme qu'aucun contrôle n'est disponible à ce stade.
- Ce point n'est toutefois pas bloquant car ces contrôles sont liés au formalisme d'archivage spécifique CINES basé sur le Dublin Core. Le scénario privilégié actuellement est la conversion en amont des SIP vers du SEDA 2.1. Il faut réétudier la question des relations au regard des nouvelles possibilités offertes par le SEDA 2.1.

Ent13 : Vérification des métadonnées d'évaluation et de communicabilité (formalisme DublinCore)

- Concernant les métadonnées: AppraisalRule et AccessRule, le besoin est couvert car les SIP DublinCore du CINES seront mappés vers du SEDA 2.1.
- Elles seront contrôlées via le PA ou le PUA déjà déclarés dans le SIP.

Ent14 : Validation Profil d'archivage Dublin Core CINES

- Le besoin est couvert via la rédaction de PUA indiquant les contraintes à appliquer aux métadonnées DublinCore déjà ajoutées en extension du SEDA 2.1.

Ent15 - Validation Profil d'archivage SEDA CINES

- Il est possible de substituer le xsd du schéma SEDA-Vitam par un xsd SEDA-CINES.
- Toutefois, cette fonctionnalité n'a pas été testée dans le cadre du POC.

Ent16 - Validation Profil d'archivage SEDA Service versant

- Les contrôles réalisés par VITAM sur ce point sont similaires à ceux réalisés par PAC.

Ent17 - Validation schematron du bordereau de versement

- VITAM ne fait pas de contrôle en entrée de type « schematron » pour contrôler la cohérence des éléments déclarés dans les métadonnées de gestion.
- Cette gestion des incohérences (notamment sur les DUA ou la communicabilité) se fait au moment de l'accès à la donnée.
- Compte tenu d'une part de la complexité du schematron SEDA 1.0 du CINES et de sa maintenance,

et d'autre part, de la structure plus complexe du SEDA 2.1, l'utilisation d'un schematron pour contrôler le SEDA 2.1 n'est pas envisagée à ce stade.

Ent18 - Vérification de la taille des fichiers

- Le contrôle effectué dans VITAM est le même que celui réalisé dans PAC, à ce ci près que la balise Size est obligatoire en SEDA 2.1 alors qu'elle était facultative en SEDA 1.0 et inexistante en SIP Dublin Core CINES.
- L'outil de conversion des SIP CINES en SEDA 2.1 devra donc intégrer la génération de cette métadonnée de manière automatique.

Ent19 - Vérification de la présence d'au moins un fichier dans le répertoire "id-archive" du SIP

- Vitam contrôle la présence du fichier obligatoire manifest.xml
- Contrairement au fonctionnement actuel de PAC, Vitam accepte un SIP même si le répertoire "content" est vide ; l'exemple type étant la déclaration d'un arbre de positionnement ou d'un plan de classement.

Ent20 - Validité de la convention d'archivage

- Dans VITAM, la métadonnée ArchivalAgreement sert à renseigner le contrat d'entrée, tandis que dans PAC SEDA 1.0, elle sert à renseigner l'identifiant de la convention d'archivage, or le contrat d'entrée n'est pas l'équivalent de la convention.
- Il faudra prévoir de rajouter dans l'interface CINES de Vitam un paramètre supplémentaire pour lier la convention d'archivage avec les contrats d'entrée.

Ent21 - Validité du format des fichiers

- Aucun contrôle sur la validité de format des objets présents dans le paquet n'est pour l'instant réalisé par VITAM.
- Ce point est à l'étude pour rapprocher des réflexions autour de la notion de classe de service et de la possibilité de rajouter des griffons externes pour appeler le validateur de formats du CINES.

Suite aux ateliers avec Vitam enDécembre 2018 :

- Une modification de Vitam a été réalisée dans la V2 livrée le 18/02 concernant la possibilité d'utiliser des griffons.
- L'objectif d'un de ces griffons est de contrôler la validité du format des fichiers en cours d'archivage.
- Il est prévu également que ce contrôle puisse être réalisé en amont de l'entrée d'une archive ce qui correspond au cas d'utilisation de PAC.

Ent22 - Vérification de signature électronique

- Aucune vérification de signature électronique n'est effectuée à l'entrée par VITAM.
- Elle doit être vérifiée en amont, et les métadonnées peuvent être conservées dans le bloc Signature du Content de l'AU.

Ent23 – Contrôles BIR (Bibliothèque d'Informations de Représentation)

- La notion de BIR n'existe pas dans Vitam, il faudra donc l'intégrer sous la forme d'un griffon Vitam.

Ent24 - Génération de l'AIP

- Lors de la réception d'un SIP, Vitam réalise une transformation du SIP en AIP et génère également un bordereau d'état d'importation notifié au service versant.

Divergences

- PAC implémente la version simple de l'OAIS : 1 SIP = 1 AIP.
- VITAM permet de créer plusieurs AIP à partir d'un SIP ou de créer un AIP à partir de plusieurs SIP.
- Il serait donc intéressant de repenser les pratiques du CINES au regard de ces nouvelles possibilités.

Ent25 - Génération d'un PID

- Un griffon est prévu pour permettre à Vitam d'intégrer l'identifiant ARK, (ou autre système d'identifiant unique), dans les bordereaux de versement

Ent26 – Calcul d'empreinte

- Les calculs d'empreintes réalisés par VITAM sont plus diversifiés que ceux réalisés dans PAC.
- Vitam peut contrôler des algorithmes en entrée MD5, SHA-1, SHA-256, SHA-512, alors que PAC se limite à l'algorithme SHA-256.

Ent27 - Indexation des métadonnées d'archivage dans un moteur d'indexation

- Le moteur d'indexation utilisé par Vitam est le même que celui utilisé dans PAC.

Ent28 - Indexation des métadonnées métier

- Dans le formalisme d'archivage SEDA 1.0, il est possible d'ajouter, dans le bordereau de versement, des métadonnées métier en tant que métadonnées externes au SEDA.
  → Indexation possible dans VITAM, en déclarant ces nouvelles métadonnées dans l'ontologie et éventuellement dans un PA.
- Dans le formalisme d'archivage DublinCore CINES, les métadonnées métier sont présentes sous la forme d'un fichier distinct joint au paquet d'archives (dans le sous-répertoire DESC).
  → Réflexions à avoir sur la manière de prendre en compte ce fichier dans VITAM.

Ent29/30/31 – Traitement OCR, Indexation et Connexion avec le HPC

- Le traitement OCR est possible à l'aide de l'implantation dans Vitam d'un plugin spécifique.
- Les tests réalisés démontrent qu'il est possible de réaliser la génération d'OCR de données sur des fichiers désignés.
- L'indexation des résultats de traitements OCR peut se faire en implémentant un plugin/greffon qu'il faut exécuter au niveau de l'archive.
- Ce point est à rapprocher des réflexions autour de la notion de classe de service et de la possibilité de rajouter des greffons externes dans le workflow d'archivage.
- Vitam propose dans sa dernière release l'ajout de scénarii de préservation avec l'import possible de « griffons » notamment celui qui permet de faire de la génération d'OCR de document à l'aide de l' applicatif externe « Tesseract ».
- Actuellement, l'utilisation de ces scénarii n'est possible que dans le cadre de la préservation des archives, mais la feuille de route de Vitam prévoit de les rendre paramétrables dans le processus

d'ingestion d'une archive.

Ent32 - Paramétrage des contrats d'entrée

- Plusieurs éléments sont actuellement paramétrables dans le contrat d'entrée de VITAM, qui recommande de créer autant de contrat d'entrée qu'il y a de profil d'archivage, chaque contrat d'entrée faisant référence au profil d'archivage et à un nœud de rattachement dans le plan de classement.
- Toutefois, il n'est pas possible, actuellement, de choisir un workflow différent en fonction du contrat d'entrée.
- A noter qu'on ne peut pas rattacher les profils d'unité archivistique (PUA) à un contrat d'entrée (CE).
- Cette notion de contrat d'entrée se rapproche de l'idée de classe de service au CINES.
- Toutefois, ils manquent un certain nombre d'éléments avant de pouvoir paramétrer l'intégralité du workflow d'entrée.
- Ce point doit intégrer la réflexion en cours sur la possibilité de rajouter des greffons externes dans le workflow d'archivage et de les lier via le contrat d'entrée.

Ent33/34 – Prise en compte des métadonnées au format Dublin Core et SEDA 1.0

- Le CINES qui utilise actuellement un formalisme de SIP spécifique dont le bordereau de versement repose en partie sur du Dublin Core et le formalisme du SEDA 1.0, n'envisage pas d'abandonner ces deux formalismes en entrée pour s'adapter à Vitam qui n'accepte que des paquets conformes à la norme SEDA 2.1.
- Le CINES prévoit le développement d'un outil de conversion de SIP conformes à VITAM (SEDA 2.1) à partir de SIP CINES en DublinCore et SEDA 1.0.

Gestion des archives existantes

GAE01 - Administration des règles de gestion

- Vitam propose un panel complet de fonctionnalités de mise à jour des règles de gestion, incluant le choix de l'héritage des règles, le pré calcul des dates d'échéance...
- A noter :
    - Le principe de prédominance pour VITAM des règles définies dans les DescriptiveMetadata des AU, par rapport à celles définies globalement dans les ManagementMetadata.
    - L'héritage ou non héritage des règles n'est pas écrit en dur dans la base de données pour chaque niveau de description.
    - Des réflexions sont en cours pour tenir compte des règles héritées lors des requêtes d'accès.

GAE02 - Mise à jour des métadonnées

- Vitam offre la possibilité de mettre à jour les métadonnées des unités d'archives et des groupes d'objets directement dans la base de données via une API (application programming interface). Cette mise à jour est possible dans PAC par l'importation d'un nouveau SIP.
- Cette mise à jour remplace l'ancienne valeur par la nouvelle, et trace cette opération dans le journal du cycle de vie, y compris la valeur écrasée. Contrairement à Vitam, PAC met à jour la valeur tout en conservant l'ancienne.
- Les deux systèmes alertent le service versant de toute modification de base de données par une

notification.

- La méthode proposée par VITAM apparaît plus souple mais une attention particulière doit être portée sur l'attribution des droits dans les contrats d'accès pour garder une certaine maîtrise du contenu du SAE.

GAE03 - Gestion des éliminations

- La procédure d'élimination se fait en deux phases : Analyse et Action.
- Cette fonction, majoritairement manuelle dans PAC, est entièrement intégrée dans VITAM, ce qui est un point positif.
- Le besoin est bien couvert : réalisation de contrôles sur les données, gestion des incohérences entre les différents niveaux de description, différents statuts de sortie.
- Quelques pistes d'améliorations :
    - Désynchronisation des contrôles sur les données entre les 2 phases
    - Incohérence du statut final en cas de non élimination
    - Difficulté d'identification des AU en conflit dans le rapport final

GAE04 – Transfert vers un autre SAE

- Cette fonctionnalité n'étant pas disponible dans la version testée (prévue en R10) elle n'a pu être évaluée.

Accès

Acc01 - Paramétrage des contrats d'accès

- Plusieurs éléments sont paramétrables dans le contrat d'accès de VITAM.
- La fonctionnalité répond correctement au besoin.

Acc02 - Recherche avancée ou libre dans le catalogue des métadonnées

- ElasticSearch propose deux manières différentes d'indexation pour offrir des fonctionnalités de recherche plein texte ou exacte selon le type de métadonnées.
- L'usage des filtres de recherche permet d'affiner les résultats.
- La fonctionnalité répond correctement au besoin.

Acc03 - Commande de DIP

- La possibilité de commander des DIP dépend des autorisations données dans le contrat d'accès.
- La fonctionnalité répond correctement au besoin.

Acc04 - Génération de DIP

- La fonctionnalité de génération de DIP est conforme aux exigences attendues pour le périmètre de PAC.

Acc05 - Stockage temporaire pour un accès rapide

- VITAM permet d'ajouter des griffons qui implémenteraient la récupération de fichiers versés et les résultats de traitements HPC afin de les transmettre vers une zone de diffusion ayant un accès plus rapide.
- Les détails sont à définir mais cela est tout à fait possible, à condition de configurer le mode d'accès

à cette zone de diffusion.

Acc06 - Consultation d'un relevé de valeur probante

- La fonctionnalité répond correctement au besoin.

Acc07 - Gestion des relations entre AIP

- Les relations entre AIP, telles que définies dans PAC, ne sont pas prises en compte dans VITAM.
- Toutefois, au vu de notre étude, ce point n'est pas complètement bloquant.
- VITAM permettant de dissocier l'unité des SIP de celles des AIP, il faut repenser la pratique actuelle des relations entre AIP en utilisant cette nouvelle possibilité.

Préservation

Pre01 - BIRv2

- La Bibliothèque d'Information et de Représentation (BIR) de PAC est utilisée pour archiver l'ensemble des documents transverses au SAE, qu'il s'agisse de la documentation du système (spécifications du SAE, spécifications des formats de fichiers archivés, etc.) ou des documents relatifs aux différents projets d'archivage (conventions, profils d'archivage, etc.).
- Il n'existe pas d'espace similaire prévu dans VITAM mais rien n'empêche d'en créer un.

Pre02/03 – Gestion de la liste et de la validation des formats d'archivage

- La validation de formats a été introduite avec la notion de « griffon » qui permet l'ajout de traitements externes à VITAM effectués par des outils tiers.
- Vitam propose notamment le paramétrage de scénarii de préservation qui s'appuient sur des applicatifs externes (Jhove, Exiftool, ImageMagick, Vera PDF …).
- Cette fonctionnalité n'a pas pu être testée directement car livrée en R9 mais plusieurs échanges ont eu lieu pour s'assurer qu'elle réponde correctement aux besoins du CINES, relativement précis sur le sujet.
- Actuellement, l'utilisation de ces scénarii n'est possible que dans le cadre de la préservation des archives, mais la feuille de route de Vitam prévoit de les rendre paramétrables dans le processus d'ingestion d'une archive.

Pre04 - Réalisation des audits

- Vitam propose la réalisation de trois types d'audit au niveau de la gestion des archives :
- Vérification du nombre de copies,
- Intégrité des objets sur les offres de stockage par rapport à la base
- Ou aux fichiers sécurisés.
- Cette fonctionnalité est conforme au niveau d'exigences attendu dans PAC.

Pre05 - Administration des stratégies de conservation

- Cette fonctionnalité introduit la notion de « griffon » et permet l'ajout de traitements externes à VITAM effectués par des outils tiers.
- Cette fonctionnalité n'a pas pu être testée directement car livrée en R9 mais plusieurs échanges ont eu lieu pour s'assurer qu'elle réponde correctement aux besoins du CINES, relativement précis sur

le sujet (cf. fonctions Pre04, Ent29 et Ent30 notamment).

Administration fonctionnelle

AFc01 – Monitoring Système (Machine)

- Vitam ne propose pas de monitoring système pour les machines mais il est tout à fait possible de connecter les machines utilisées à une application comme Shinken (utilisé au CINES) afin de réaliser la surveillance système et réseau.

AFc02 – Monitoring Système (Flux)

- Vitam propose un monitoring Système pour les flux (Consul et Kibana Log) qui est adapté aux exigences du CINES.

AFc03 – Monitoring métier

- Vitam propose un monitoring métier (Kibana Data), essentiellement à destination du service d'archives, adapté aux exigences du CINES.

AFc04– Journalisation

- La gestion de la journalisation par VITAM correspond aux besoins du CINES et aux exigences de la norme NF Z42-013. Les fonctionnalités offertes sont sensiblement différentes de celles présentes dans le SAE du CINES.
- Il est à noter que la gestion des opérations est directement indexée dans le moteur de VITAM, ce qui en rend l'exploitation aisée.

AFc09 – Gestion des habilitations

- Vitam propose la possibilité de gérer les habilitations d'une manière beaucoup plus fine que celle réalisée dans PAC.

AFc11 – Fourniture de statistiques

- Vitam propose des statistiques applicatives et métier (à l'aide des tableaux de bord de l'application Kibana) qui sont adaptées aux exigences du CINES.

Administration technique

ATe01 – Gestion des habilitations

- La connexion d'accès au composant externe de Vitam se fait par authentification de certificat applicatif (Accès / Entrées dans Vitam) ce qui assure une sécurité et un contrôle renforcés des connexions des applications externes.
- Ces certificats applicatifs sont liés à un profil de sécurité déclaré et configurable dans Vitam.
- Le choix des permissions associées à l'accès aux fonctionnalités de Vitam se fait via ce profil.

ATe02 – Monitoring système (service)

- Vitam propose plusieurs outils d'exploitation pour sa solution.
- Un annuaire de service « Consul » permet notamment de connaitre l'état des différents services pour s'assurer du bon fonctionnement d'un composant.

Stockage et Gestion des données

SGD01 – Offres de stockage

- Vitam propose deux types d'offre de stockage :
  - Une offre « Filesystem » qui permet de stocker des données sur un système de fichiers.
  - Une offre « Swift » qui permet de stocker des données sur un stockage objet implémentant l'API Swift.
- Aucune offre froide n'est disponible actuellement dans Vitam. La possibilité de faire du stockage sur bande est prévue dans la feuille de route de Vitam (Release 10).
- Le CINES et l'équipe Vitam réaliseront une expérimentation sur l'implémentation du stockage sur bandes.

SGD02 - Gestion des copies

- Vitam possède un composant de stratégie de stockage qui permet de déclarer autant d'offres de stockage que souhaitées. Chaque offre de stockage déclarée dans la stratégie correspond à une copie.
- Tout en n'introduisant pas d'incompatibilité, il est à noter que cette gestion de copie est différente de celle effectuée dans PAC. En effet, dans une offre de stockage PAC, on déclare le nombre de copies requises.

SGD03 - Copie distante

- Vitam propose la possibilité d'une installation multi-sites et donc d'offres de stockage multi-sites associées.

SGD04 – Reprise sur erreur d'écritures

- Dans le cas d'une erreur technique lors d'un versement, le processus est mis à l'état PAUSE.
- La reprise du versement n'est pas automatique mais est réalisable par deux types de requête :
  - RESUME, reprise totale du versement
  - NEXT, reprise pas à pas du versement

SGD05 – Zones obligatoires

- Dans Vitam, l'écriture sur l'ensemble des offres définies dans la stratégie de stockage est obligatoire.
- La notion de rendre une zone obligatoire ou non dans Vitam n'existe pas.

SGD06 – Emplacement de stockage

- Vitam possède une architecture multi-tenant qui permet la gestion séparée des informations stockées pour chaque client.
- Il serait donc possible de dédier chaque projet d'archive à un emplacement de stockage spécifique (logique) en dédiant à chaque service versant du CINES un tenant spécifique.

SGD07 – Gestion des priorités

- La stratégie de stockage dans Vitam permet de définir une offre dite « primaire » utilisée prioritairement pour tout ce qui est gestion des accès, communication …

SGD08 – Gestion des supports disques (Migration physique)

- Cette fonctionnalité n'est actuellement pas implémentée dans Vitam mais fait partie de leur roadmap.

SGD09 – Archivage / tape (Interface SCSI)

- Cette fonctionnalité n'est actuellement pas implémentée dans Vitam et reste en cours d'étude. Prévue pour la release 10 de VITAM.

SGD10 – Gestion des supports bandes

- Cette fonctionnalité n'est actuellement pas implémentée dans Vitam et reste en cours d'étude. Prévue pour la release 10 de VITAM.

SGD11 – Contrôle intégrité

- Un audit sur la cohérence et l'intégrité des données stockées sur une offre est réalisable dans Vitam.
- Il est possible alors de resynchroniser les données entre deux offres de stockage.

Transverse

TRA01 - Maintenance et amélioration continue (MAC)

- Maintenance opérée par le SDSI du Ministère de la Culture
- Devra être active fin 2019 et prévoir un tuilage avec l'équipe actuelle pour la transmission des connaissances.
- Budget annuel de 1,9 M d'€ payé par les 3 ministères porteurs, 12 ETP dont 5 internes (pour le pilotage et la conception) et 7 externes (pour le dév). Sans les 2 experts formats (proposition non validée par le CODIR !).
- Club utilisateurs :
  - pour échanger sur l'outil et élaborer en commun des demandes relatives à la maintenance et/ ou à l'amélioration continue du produit, avec un prisme privilégié sur l'archivage intermédiaire, pour assurer la réutilisation
  - paiement d'un forfait de support (ticket déterminé selon la taille de l'entité et les prix du marché de développement et support).
  - désignation d'un 4ème acteur pour être le représentant des utilos dans les instances de pilotage (en plus des 3 ministères porteurs).
- Décisions prises par les instances de pilotage collégiales (priorité des corrections, plan d'évolution, etc.) par consensus par les 3 ministères initiateurs du programme, après consultation du 4e acteur, et dans la mesure du possible, en consensus avec ce dernier. Si le 4e acteur estime que des intérêts majeurs des membres du club utilisateurs n'ont pas été pris en considération, il pourra solliciter l'avis de la DINSIC et du CIAF.

TRA02 - Utilisation de SHERPA (Générateur de profils d'archivage)

Remarques générales: outil peu intuitif; la documentation produite en html est moins lisible que celle produite par Agape pour le SEDA 1.0 ce qui entraîne des difficultés pour la relecture d'un profil.

SHERPA ne produit que du SEDA 2.0, mais Vitam demande du 2.1. Il faut donc modifier directement le fichier RNG produit par SHERPA pour le rendre compatible au SEDA 2.1. D'autre part, il est impossible de

recharger un profil modifié dans l'application.

Inconvénients principaux:

- plus de temps pour élaborer un profil
- risque d'erreurs
- complexité pour mettre à jour un profil

D'autre part, il faut penser à l'impact sur les futurs utilisateurs. Doutes sur leur capacité à pouvoir utiliser l'outil et à apporter les modifications nécessaires pour rendre les profils compatibles au 2.1. Faut-il envisager que désormais, le CINES prendrait en charge la rédaction des profils?

[VITAM-PAC] Récapitulatif

Récapitulatif de l'étude fonctionnelle comparative	
Version de VITAM	R7 de Vitam, livrée en milieu d'année 2018
Période	6 Mois (Juin 2018 – Décembre 2019)
Equipe Projet CINES	13 personnes
Equipe Projet CROUS	2 personnes
Equipe Projet ATOS	4 personnes
Méthode utilisée	Agile – Kanban
Nombre d'axes d'étude	9
Nombre de fonctionnalités étudiées	85
Couverture fonctionnelle de VITAM / PAC	85,3 %