



## Gestion de la préservation

Date	Version
24/04/2019	1.0 (Release 10)

### État du document

En projet     Vérifié     Validé

## Maîtrise du document

Responsabilité	Nom	Entité	Date
Rédaction	MVI	Équipe Vitam	10/12/2018
Vérification	Équipe	Équipe Vitam	
Validation	EVA	Équipe Vitam	

## Suivi des modifications

Version	Date	Auteur	Modifications
0.1	10/12/2018	MVI	Initialisation
0.2	22/02/2019	JSL	Relecture et corrections
0.3	28/02/2019	MVI	Corrections
0.4	05/03/19	Équipe Vitam	Relecture et corrections
0.5	18/04/19	MVI	Corrections. Points d'attention : le présent document nécessitera les compléments suivants dans une prochaine version : <ul style="list-style-type: none"> <li>• explicitation des mécanismes mis en place dans le cadre d'une extraction de métadonnées au sein de la solution logicielle Vitam ;</li> <li>• améliorations des explications relatives à l'audit de cohérence et à l'audit correctif.</li> </ul>
0.6	19/04/19	MVI	Relecture
1.0	24/04/2019	MRE	Finalisation du document pour publication de la Release 10.

## Documents de référence

Document	Date de la version	Remarques
NF Z44022 – MEDONA - Modélisation des données pour l'archivage	18/01/2014	
Standard d'échange de données pour l'archivage – SEDA – v. 2.1	10/07/05	
Vitam - Structuration des <i>Submission Information Package</i> (SIP) – v. 8.0.	24/04/2019	

## **Licence**

La solution logicielle VITAM est publiée sous la licence CeCILL 2.1 ; la documentation associée (comprenant le présent document) est publiée sous Licence Ouverte V2.0.

## Table des matières

1. Résumé.....	8
1.1. Présentation du programme Vitam.....	8
1.2. Présentation de la problématique.....	9
1.3. Présentation du document.....	10
2. Administration de la préservation numérique.....	11
2.1. Référentiel des formats.....	11
2.1.1. Définition.....	11
2.1.2. Formalisation.....	12
2.1.2.1. Dans le registre PRONOM.....	12
2.1.2.2. Dans la solution logicielle Vitam.....	14
2.1.3. Mécanismes mis en œuvre dans la solution logicielle Vitam.....	15
2.1.3.1. Import.....	15
2.1.3.2. Ré-import et mise à jour.....	16
2.1.4. Conseils de mise en œuvre.....	17
2.1.4.1. À quoi sert le référentiel des formats ?.....	17
2.1.4.2. Quand et comment créer un référentiel des formats ?.....	18
2.1.4.3. Quand et comment ré-importer et mettre à jour le référentiel des formats ?.....	18
2.1.4.4. Comment procéder lors d’une mise à jour du référentiel des formats ?.....	19
2.1.4.5. Comment utiliser le référentiel des formats ?.....	20
2.2. Outils de préservation (griffons).....	21
2.2.1. Définition.....	21
2.2.2. Formalisation.....	22
2.2.2.1. Dans un fichier JSON.....	22
2.2.2.2. Dans la solution logicielle Vitam.....	22
2.2.3. Mécanismes mis en œuvre dans la solution logicielle Vitam.....	23
2.2.3.1. Implémentation des griffons.....	23
2.2.3.2. Gestion du référentiel des griffons.....	24
2.2.3.2.1. Import.....	24
2.2.3.2.2. Ré-import et mise à jour.....	24
2.2.4. Conseils de mise en œuvre.....	25
2.2.4.1. À quoi sert le référentiel des griffons ?.....	25
2.2.4.2. Quand et comment créer un référentiel des griffons ?.....	25
2.2.4.2.1. Au moment de l’installation de la solution logicielle Vitam.....	25
2.2.4.2.2. Après installation de la solution logicielle Vitam.....	26
2.2.4.3. Comment nommer un exécutable ?.....	27
2.2.4.4. Quand et comment ré-importer et mettre à jour le référentiel des griffons ?.....	27

2.2.4.5. Comment procéder lors d'une mise à jour du référentiel des griffons ?.....	28
2.2.4.6. Quand et comment supprimer un griffon ?.....	28
2.2.4.7. Comment utiliser le référentiel des griffons ?.....	29
2.3. Scénarios de préservation.....	29
2.3.1. Définition.....	29
2.3.2. Formalisation.....	31
2.3.2.1. Dans un fichier JSON.....	31
2.3.2.2. Dans la solution logicielle Vitam.....	33
2.3.3. Mécanismes mis en œuvre dans la solution logicielle Vitam.....	35
2.3.3.1. Import.....	35
2.3.3.2. Ré-import et mise à jour.....	35
2.3.4. Conseils de mise en œuvre.....	36
2.3.4.1. À quoi sert le référentiel des scénarios de préservation ?.....	36
2.3.4.2. Quand et comment créer un référentiel des scénarios de préservation ?.....	36
2.3.4.3. Quand et comment ré-importer et mettre à jour le référentiel des scénarios de préservation ?.....	37
2.3.4.4. Quand et comment créer un scénario de préservation ?.....	38
2.3.4.4.1. Réflexion autour des besoins en termes de préservation.....	38
2.3.4.4.2. Rédaction d'un scénario de préservation.....	38
2.3.4.5. Quand et comment supprimer un scénario de préservation ?.....	41
2.3.4.6. Comment élaborer et structurer un référentiel des scénarios de préservation ?.....	41
2.3.4.7. Comment utiliser le référentiel des scénarios de préservation ?.....	43
3. Entrées et préservation.....	45
3.1. Définition.....	45
3.2. Mécanismes mis en œuvre par la solution logicielle Vitam.....	45
3.2.1. Filtres sur les entrées.....	45
3.2.2. Contrôles sur les entrées.....	46
3.2.2.1. Service d'identification.....	46
3.2.2.2. Service supplémentaire.....	47
3.3. Conseils de mise en œuvre.....	48
3.3.1. Quand et comment ajouter des contrôles sur les entrées ?.....	48
3.3.2. Comment paramétrer les filtres sur les entrées ?.....	49
4. Gestion des archives et préservation.....	51
4.1. Préservation.....	51
4.1.1. Définition.....	51
4.1.2. Mécanismes mis en œuvre dans la solution logicielle Vitam.....	52
4.1.2.1. Définition/Paramétrage de l'opération de préservation.....	52
4.1.2.2. Opération de préservation.....	52
4.1.2.3. Résultats d'une opération de préservation.....	54

4.1.3. Conseils de mise en œuvre.....	55
4.1.3.1. Quel accès à la gestion de la préservation ?.....	55
4.1.3.1.1. Gestion des droits.....	55
4.1.3.1.2. Restitution sur une IHM.....	56
4.1.3.2. Comment utiliser et gérer la préservation ?.....	56
4.2. Audit.....	62
4.2.1. Audit d’existence et d’intégrité des objets.....	62
4.2.1.1. Définition.....	62
4.2.1.2. Mécanismes mis en œuvre dans la solution logicielle Vitam.....	63
4.2.1.1.1. Sélection de l’objet à auditer.....	63
4.2.1.1.2. <i>Opération d’audit</i> .....	63
4.2.1.1.2. <i>Résultat d’une opération d’audit</i> .....	64
4.2.1.3. Conseils de mise en œuvre.....	64
4.2.1.3.1. <i>À quoi servent l’audit d’existence et l’audit d’intégrité des objets ?</i> .....	64
4.2.1.3.2. <i>Quand et comment lancer un audit d’existence et un audit d’intégrité des objets ?</i>	64
4.2.1.3.3. <i>Comment utiliser l’audit d’existence et d’intégrité des objets ?</i> .....	65
4.2.2. Audit de cohérence.....	67
4.2.2.1. Définition.....	67
4.2.2.2. Mécanismes mis en œuvre dans la solution logicielle Vitam.....	68
4.2.2.2.1. Sélection de l’objet à auditer.....	68
4.2.2.2.2. <i>Opération d’audit</i> .....	68
4.2.2.2.3. <i>Résultat d’une opération d’audit</i> .....	69
4.2.2.3. Conseils de mise en œuvre.....	69
4.2.2.3.1. <i>À quoi sert l’audit de cohérence ?</i> .....	69
4.2.2.3.2. <i>Quand et comment lancer un audit de cohérence ?</i> .....	70
4.2.2.3.3. <i>Comment utiliser l’audit de cohérence ?</i> .....	70
4.2.3. Audit correctif.....	71
4.2.3.1. Définition.....	71
4.2.3.2. Mécanismes mis en œuvre dans la solution logicielle Vitam.....	72
4.2.3.2.1. Sélection de l’objet à auditer.....	72
4.2.3.2.2. <i>Opération d’audit</i> .....	72
4.2.3.2.3. <i>Résultat d’une opération d’audit</i> .....	73
4.2.3.3. Conseils de mise en œuvre.....	73
4.2.3.3.1. <i>À quoi sert l’audit correctif ?</i> .....	73
4.2.3.3.2. <i>Quand et comment lancer un audit correctif ?</i> .....	74
4.2.3.3.3. <i>Comment utiliser l’audit correctif ?</i> .....	74
Annexe 1 : exemples de référentiels de préservation.....	76
A. Griffons.....	76
B. Scénarios de préservation.....	76

Annexe 2 : Messages d'erreur.....	78
Annexe 3 : Paramétrages des scénarios de préservation.....	81
A. ImageMagick.....	81
B. JHOVE.....	82
C. Siegfried.....	82
D. ODFvalidator.....	82
E. VeraPDF.....	82
F. LibreOffice.....	83
G. Tesseract.....	84

## 1. Résumé

Jusqu'à présent, pour la gestion, la conservation, la préservation et la consultation des archives numériques, les acteurs du secteur public étatique ont utilisé des techniques d'archivage classiques, adaptées aux volumes limités dont la prise en charge leur était proposée. Cette situation évolue désormais rapidement et les acteurs du secteur public étatique doivent se mettre en capacité de traiter les volumes croissants d'archives numériques qui doivent être archivés, grâce à un saut technologique.

### 1.1. Présentation du programme Vitam

Les trois ministères (Europe et Affaires étrangères, Armées et Culture), combinant légalement mission d'archivage définitif et expertise archivistique associée, ont décidé d'unir leurs efforts, sous le pilotage de la Direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC), pour faire face à ces enjeux. Ils ont décidé de lancer un programme nommé Vitam (Valeurs Immatérielles Transmises aux Archives Pour Mémoire) qui couvre plus précisément les opérations suivantes :

- la conception, la réalisation et la maintenance mutualisées d'une solution logicielle d'archivage électronique de type back-office, permettant la prise en charge, le traitement, la conservation et l'accès aux volumes croissants d'archives (projet de solution logicielle Vitam) ;
- l'intégration par chacun des trois ministères porteurs du Programme de la solution logicielle dans sa plate-forme d'archivage. Ceci implique l'adaptation ou le remplacement des applications métiers existantes des services d'archives pour unifier la gestion et l'accès aux archives, la reprise des données archivées depuis le début des années 1980, la réalisation d'interfaces entre les applications productrices d'archives et la plate-forme d'archivage (projets SAPHIR au MEAE, ADAMANT au MC et ArchiPél au MA) ;
- le développement, par un maximum d'acteurs de la sphère publique, de politiques et de plates-formes d'archivage utilisant la solution logicielle (projet Ad-Essor).

La solution logicielle Vitam est développée en logiciel libre et recourt aux technologies innovantes du Big Data, seules à même de relever le défi de l'archivage du nombre d'objets numériques qui seront produits ces prochaines années par les administrations de l'État. Afin de s'assurer de la qualité du logiciel livré et de limiter les dérives calendaires de réalisation, le projet est mené selon une conduite de projet Agile. Cette méthode dite « itérative », « incrémentale » et « adaptative » opère par successions de cycles réguliers et fréquents de développements-tests-corrections-intégration. Elle associe les utilisateurs tout au long des développements en leur faisant tester les éléments logiciels produits et surtout en leur demandant un avis sur la qualité des résultats obtenus. Ces contrôles réguliers permettent d'éviter de mauvaises surprises lors de la livraison finale de la solution logicielle en corrigeant au fur et à mesure d'éventuels dysfonctionnements.

Le programme Vitam bénéficie du soutien du Commissariat général à l'investissement dans le cadre de l'action : « Transition numérique de l'État et modernisation de l'action publique » du Programme



d'investissement d'avenir. Il a été lancé officiellement le 9 mars 2015, suite à la signature de deux conventions, la première entre les ministères porteurs et les services du Premier ministre, pilote du programme au travers de la DINSIC, et la seconde entre les services du Premier ministre et la Caisse des dépôts et consignations, relative à la gestion des crédits attribués au titre du Programme d'investissements d'avenir.

## 1.2. Présentation de la problématique

La préservation numérique consiste en un ensemble de processus visant à maintenir un accès aux objets numériques sur le très long terme. Elle implique d'assurer un **moyen de représenter ces objets** dans le temps, de vérifier la **disponibilité** et l'**intégrité** des objets dans la solution logicielle Vitam.

Les processus, permettant de garantir qu'un objet est en mesure de traverser le temps durant tout son cycle de vie tout en préservant son intégrité, sont les suivants :

- l'identification de format : processus qui consiste à définir précisément le format d'un fichier numérique, qu'il s'agisse d'un conteneur<sup>1</sup> ou non ;
- la validation de format : processus qui consiste à vérifier qu'un fichier respecte les spécifications publiées de son format, en termes de structure comme de syntaxe ;
- la conversion de format : opération qui consiste à convertir un objet dans un format différent de celui dans lequel il est encodé. Elle doit préserver la fidélité du document (Source : NF Z 42013) ;
- l'extraction de métadonnées : opération qui consiste à extraire d'un fichier tout ou partie des informations ou métadonnées qu'il contient afin de les exploiter en dehors du fichier, dans une base de données, ou de les réintégrer dans celui-ci après conversion.

Pour répondre à cette problématique et à cet objectif, la solution logicielle Vitam met à disposition les outils et services suivants :

- un traitement de masse sur les objets numériques pour :
  - identifier ou ré-identifier leur format suite à des mises à jour du référentiel des formats (identification de format),
  - analyser des objets non seulement pour vérifier leur conformité par rapport aux spécifications de leur format (validation de format) mais aussi pour évaluer leur qualité par rapport à des critères émis pour des besoins de conservation (ex : vérifier la taille d'un fichier image en termes de pixels) ;
  - générer de nouveaux objets numériques, non seulement à des fins de conversion de format pour cause de format obsolète, mais aussi à d'autres fins (ex : génération de vignettes, d'objets de diffusion) ;
  - extraire tout ou partie des métadonnées contenues dans les objets numériques ;
- une administration de ce précédent service au moyen des outils suivants, qui permettent de paramétrer et d'effectuer ces traitements :

---

<sup>1</sup> Un format conteneur est un format de fichier pouvant contenir divers type de données (ex : fichiers ZIP, fichiers vidéo au format MPEG ou AVI).

- griffons ;
- scénarios de préservation ;
- référentiel des formats ;
- des contrôles en entrée du système en vue d’identifier les formats des objets numériques ou de contraindre les objets pouvant entrer dans la solution logicielle Vitam ;
- des services de sécurisation des journaux, d’audit et de génération de relevé de valeur probante, en vue de garantir l’intégrité, la conservation des objets numériques et leur accès au sein de la solution logicielle Vitam.

### 1.3. Présentation du document

Le document présente les fonctionnalités associées à la gestion et à la réalisation d’opérations de préservation numérique dans la solution logicielle Vitam.

Il s’articule autour des axes suivants :

- une présentation des différents référentiels associés à la préservation numérique : formats, griffons, scénarios de préservation, et de la manière dont ils doivent être formalisés en JSON ou en XML, ainsi que dans le modèle de données de la solution logicielle Vitam ;
- une présentation des mécanismes mis en œuvre dans la solution logicielle Vitam pour gérer et réaliser des opérations de préservation numérique ;
- des recommandations aux utilisateurs de la solution logicielle Vitam sur la manière d’utiliser les fonctionnalités associées à ces outils de préservation numérique.

Le présent document décrit les fonctionnalités qui sont offertes par la deuxième version de production de la solution logicielle Vitam au terme de la *release 10* (avril 2019). Il a vocation à être amendé, complété et enrichi au fur et à mesure de la réalisation de la solution logicielle Vitam et des retours et commentaires formulés par les ministères porteurs et les partenaires du programme.

## 2. Administration de la préservation numérique

La solution logicielle Vitam s'appuie sur **trois référentiels** pour administrer, gérer et paramétrer les actions de préservation :

- le référentiel des **formats**,
- le référentiel des **griffons**, ou outils de préservation,
- le référentiel des **scénarios de préservation**.

### 2.1. Référentiel des formats

#### 2.1.1. Définition

La préservation numérique implique d'effectuer des traitements sur des objets numériques en s'appuyant sur une identification fiable de ces derniers. Cette identification est assurée, référencée, harmonisée et normalisée au moyen d'un référentiel des formats, unique dans le système.

Par nature, un référentiel des formats identifie un ensemble de formats, en leur associant les informations nécessaires à leur identification, et fournit les moyens de contrôler les objets conservés dans la solution logicielle Vitam.

La solution logicielle Vitam utilise comme référentiel des formats le **registre PRONOM**, développé en mars 2002 par The National Archives (UK) pour faciliter l'accès et la préservation sur le long terme aux archives numériques qu'elle conservait. Il s'agit de la principale ressource disponible en ligne sur les formats de fichiers connus ainsi que sur les produits logiciels capables de les créer et de les lire – y compris leur cycle de vie et leurs spécifications techniques. Mis en ligne à partir de février 2004, ce registre continue de s'enrichir et contient en septembre 2018 plus de 2000 entrées. Dans le registre PRONOM, chaque format de fichier dispose d'un identifiant unique, le **Pronom Unique IDentifier (ou PUID)** et est décrit avec les informations suivantes :

- un résumé des informations essentielles disponibles :
  - nom du format ;
  - version du format ;
  - identifiants (PUID, Type MIME) ;
  - catégorie ;
  - description ;
  - formats de fichiers associés ;
  - date de mise à disposition ;
  - dates de création et de mise à jour de la notice ;
- la documentation disponible, notamment si le format a fait l'objet d'une normalisation ;
- les moyens d'identification disponibles :
  - extensions ;

- nombres magiques ou signatures<sup>2</sup> ;
- les méthodes de compression utilisées, si elles existent ;
- les méthodes d’encodage des caractères, si nécessaire ;
- les droits de propriété intellectuelle associés ;
- les fichiers de référence ;
- les propriétés du format.

La solution logicielle Vitam intègre une version simplifiée du registre PRONOM en tant que référentiel des formats, pour 2 usages :

- l’un destin à la gestion des formats au sein de la solution logicielle Vitam. Ce référentiel n’enregistre pas l’ensemble des informations contenus dans le registre PRONOM ;
- l’autre intégré dans l’outil d’identification de formats utilisé par la solution logicielle Vitam. Ce dernier utilise et interprète l’ensemble des informations contenues dans le registre PRONOM.

## 2.1.2. Formalisation

### 2.1.2.1. Dans le registre PRONOM

La version simplifiée du registre PRONOM que la solution logicielle Vitam utilise prend la forme d’un fichier XML, appelé « fichier de signatures » (ou « DROID\_SignatureFile\_VXX.xml »).

Pour un format donné, ce fichier :

- déclare ses caractéristiques, en particulier les informations permettant de l’identifier, mais aussi ses relations avec d’autres formats ;
- référence les signatures internes, éléments informationnels qui permettent d’identifier un format en fonction de son train de bits<sup>3</sup>, et les associe à un ou plusieurs formats.

Le registre PRONOM inclut également des signatures externes qui ne sont pas décrites dans le fichier de signatures.

Les relations entre formats et signatures peuvent être modélisées de la manière suivante<sup>4</sup> :

De ce registre, la solution logicielle Vitam ne retient **pour son référentiel interne** que les éléments suivants :

- informations liées à la version et à la date de création du fichier de signatures, présents dans l’élément racine « FFSignatureFile » de ce dernier ;

Élément racine du fichier de signatures en date du 17 septembre 2018, version 94 :

```
<FFSignatureFile xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.nationalarchives.gov.uk/pronom/SignatureFile DROID_SignatureFile.xsd"
xmlns="http://www.nationalarchives.gov.uk/pronom/SignatureFile" DateCreated="2018-09-17T12:54:53"
```

2 Une « signature interne », ou « nombre magique » (*magic number*), consiste en une constante numérique ou un ensemble de caractères propre à un format de fichier, embarquée dans un fichier, généralement positionnée à un endroit déterminé du fichier (souvent à son début mais pas uniquement) et utilisée pour désigner le format ou le protocole à utiliser. Cette séquence n’est pas directement visible de l’utilisateur, mais elle peut être récupérée en utilisant un éditeur hexadécimal.

3 Cf. note précédente.

4 Schéma issu de BROWN Adrian, « Digital Preservation Technical Paper 1 : Automatic Format Identification Using PRONOM and DROID », The National Archives, 7 mars 2006, p.6.

Version="94">

- caractéristiques des formats, présentes dans l'élément « FileFormat ».

Caractéristiques du format dont le PUID est x-fmt/17 :

```
<FileFormat ID="44" MIMETYPE="application/vnd.ms-excel"
  Name="Microsoft Excel Template" PUID="x-fmt/17" Version="97-2003">
  <Extension>xlt</Extension>
  <HasPriorityOverFileFormatID>684</HasPriorityOverFileFormatID>
</FileFormat>
```

- signatures internes, que la solution logicielle Vitam n'enregistre pas pour son référentiel interne, mais prises en compte par les outils d'identification tels que Siegfried ou Droid pour réaliser leurs traitements. Les éléments XML du fichier de signatures sont modélisés comme suit :

Champ	Description
<b>En-tête</b>	
<FFSignatureFile>	élément racine du fichier de signatures PRONOM (obligatoire)
DateCreated	date de création du fichier PRONOM (obligatoire)
Version	version du fichier PRONOM (obligatoire)
<b>Format</b>	
FileFormat	format de fichier (facultatif et répétable)
ID	identifiant XML, correspondant à un entier (obligatoire et unique)
MIMETYPE	type MIME (facultatif)
Name	nom signifiant du format, correspondant à une chaîne de caractères ne pouvant dépasser 100 caractères et ne pouvant être vide (obligatoire)
PUID	identifiant PRONOM du format, correspondant à une chaîne de caractères unique, ne pouvant être vide et dépasser 150 caractères (obligatoire)
Version	version du format, correspondant à une chaîne de caractères ne pouvant dépasser 50 caractères (facultatif)
<Extension>	extension(s) associée(s) à un format, correspondant à une chaîne de caractères ne pouvant être vide (facultatif et répétable)
<HasPriorityOverFileFormatID>	Identifiant XML, correspondant à un identifiant de format présent dans le registre PRONOM et associé au format décrit (facultatif et répétable). Indique que ce dernier a la priorité sur ce(s) format(s), c'est-à-dire qu'il sert de format de référence et les rend obsolète.

### **Point d'attention :**

- Au terme de la release 9, la solution logicielle Vitam supporte uniquement les **versions 91 et suivantes** du référentiel PRONOM ;
- La solution logicielle Vitam n'enregistre pas les signatures internes comme externes dans son référentiel interne. Ces informations, présentes dans le fichier de signatures, sont en

revanche intégrées par les outils d'identification comme Siegfried ou Droid, en raison de leur caractère indispensable pour réaliser leurs traitements.

### 2.1.2.2. Dans la solution logicielle Vitam

Les formats sont enregistrés dans la base de données MongoDB, dans la collection « FileFormat », sous la forme d'enregistrements au format JSON.

```
{
  "_id": "aeaaaaaaho6plwab7ykalivm47lsiabza",
  "CreateDate": "2018-09-17T12:54:53.000",
  "VersionPrenom": "94",
  "UpdateDate": "2019-02-06T15:32:21.201",
  "PUID": "x-fmt/111",
  "Version": "",
  "MimeType": "text/plain",
  "Name": "Plain Text File",
  "HasPriorityOverFileFormatID": [],
  "Extension": [
    "txt"
  ],
  "Group": "",
  "Alert": false,
  "Comment": "",
  "_v": 12
}
```

Ils sont modélisés comme suit<sup>5</sup> :

Champ	Description
_id	identifiant unique dans l'ensemble du système, fourni par ce dernier (champ obligatoire)
CreateDate	date de création de la version du fichier de signatures PRONOM utilisé pour initialiser ou mettre à jour la collection, présent dans « FFSignatureFile », élément racine du fichier de signatures PRONOM (champ obligatoire).
VersionPrenom	numéro de version du fichier de signatures PRONOM utilisé pour initialiser ou mettre à jour la collection, présent dans « FFSignatureFile », élément racine du fichier de signatures (champ obligatoire).
UpdateDate	date de mise à jour de la version du fichier de signatures PRONOM au sein de la solution logicielle Vitam, fournie par cette dernière (champ obligatoire).
PUID	identifiant unique du format, fourni par le registre PRONOM (champ obligatoire). Il provient de l'attribut XML « PUID », présent dans le champ « FileFormat ». L'identifiant est composé de deux éléments séparés par un slash : <ul style="list-style-type: none"> <li>• un préfixe : « fmt » ou « x-fmt »,</li> <li>• un nombre, correspondant au numéro d'entrée du format dans le registre PRONOM.</li> </ul>

<sup>5</sup> Pour plus d'informations, consulter le document *Modèle de données*, chapitre 8, « Collection FileFormat ».

Version	version du format, fournie par le registre PRONOM (champ obligatoire, mais pouvant être vide). Il provient de l'attribut XML « Version », présent dans le champ « FileFormat ».
Name	nom du format, fourni par le registre PRONOM (champ obligatoire). Il provient de l'attribut XML « Name », présent dans le champ « FileFormat ».
MimeType	type MIME correspondant à un format de fichier, fourni par le registre PRONOM (champ obligatoire, mais pouvant être vide). Il provient de l'attribut XML « MimeType », présent dans le champ « FileFormat ».
HasPriorityOverFileFormatID	liste de PUID appartenant à des formats sur le(s)quel(s) le format décrit a la priorité et sert de format de référence, fournie par le fichier de signatures PRONOM (champ obligatoire, mais pouvant être vide). Il provient de l'élément « HasPriorityOverFileFormatID », sous-élément de « FileFormat » et s'enregistre sous forme de tableau dans la base de données MongoDB.
Group	nom d'une famille de formats (champ obligatoire, mais pouvant être vide), fourni par la solution logicielle Vitam.
Alert	alerte sur l'obsolescence du format (champ obligatoire), correspondant à un booléen dont la valeur est, par défaut, égale à « false ».
Comment	commentaires (champ obligatoire, mais pouvant être vide), fourni par la solution logicielle Vitam.
Extension	extension(s) du format, fournie par le fichier de signatures PRONOM (champ obligatoire, mais pouvant être vide). Elle provient de l'élément « Extension », sous-élément de « FileFormat » et s'enregistre sous forme de tableau dans la base de données MongoDB.
_v	version de l'enregistrement décrit, fournie par le système (champ obligatoire). <ul style="list-style-type: none"> <li>• Si l'enregistrement est égal à « 0 », il s'agit de l'enregistrement d'origine.</li> <li>• Si l'enregistrement est supérieur à « 0 », l'enregistrement a fait l'objet de 1 à n mise(s) à jour.</li> </ul>

### 2.1.3. Mécanismes mis en œuvre dans la solution logicielle Vitam

La solution logicielle intègre un référentiel des formats administrable par un utilisateur doté des droits adéquats (**administrateur fonctionnel et technique**).

Ce référentiel est multi-tenant. Il est administrable et journalisé depuis le tenant d'administration.

#### 2.1.3.1. Import

La solution logicielle Vitam intègre un référentiel des formats, **automatiquement importé lors de l'initialisation de la plate-forme**. Sa version doit être :

- identique à la version du registre PRONOM importée dans l'outil d'identification de formats utilisé par la plate-forme (Siegfried actuellement) ;
- compatible avec l'outil d'identification de formats qui, selon la version installée, ne supporte pas toutes les versions du registre PRONOM.

Ce référentiel fonctionne en mode « esclave »<sup>6</sup> puisque le PUID fourni par le registre PRONOM constitue l'identifiant de chaque format.

L'import du référentiel constitue une opération d'administration (« MASTERDATA »), tracée dans le journal des opérations du tenant d'administration de la solution logicielle Vitam<sup>7</sup>.

### 2.1.3.2. Ré-import et mise à jour

Le fichier de signatures est régulièrement mis à jour par The National Archives (UK). Il est disponible à cette adresse : <http://www.nationalarchives.gov.uk/aboutapps/pronom/droid-signature-files.htm>.

La modification du référentiel des formats est possible au moyen des API et de l'IHM standard fournie avec la solution logicielle Vitam et s'effectue par un réimport du fichier de signatures, en mode « annule et remplace ». De fait, la solution logicielle Vitam permet de :

- ajouter un nouveau format ;
- modifier les informations associées à un format ;
- supprimer un format.

La nouvelle version du référentiel des formats doit être :

- identique à la version du registre PRONOM importée dans l'outil d'identification de formats utilisé par la solution logicielle Vitam (Siegfried actuellement) ;
- compatible avec l'outil d'identification de formats qui, selon la version installée, ne supporte pas toutes les versions du registre PRONOM.

C'est pourquoi, cette action relève d'un acte d'**administration technico-fonctionnelle**.

Elle provoque la création d'une nouvelle version du référentiel. Les différentes versions du référentiel font l'objet d'une sauvegarde sur les offres de stockage utilisées par la solution logicielle Vitam.

Sauf quelques cas d'erreur, elle fait l'objet d'une journalisation dans le journal des opérations du tenant d'administration (opération d'administration de type « MASTERDATA »).

Lors de ce ré-import ou de cette mise à jour, l'opération peut aboutir aux statuts suivants :

<b>Statut</b>	<b>Motifs</b>
Succès	Opération réalisée sans rencontrer de problèmes particuliers.
Avertissement	Mise à jour d'un référentiel dont la version est identique à la version du référentiel préalablement importé.
	Mise à jour d'un référentiel dont la version est antérieure à la version du référentiel préalablement importé.

<sup>6</sup> Les PUID présents dans le fichier de signatures fourni par The National Archives (UK) sont considérés comme identifiants uniques et signifiants du référentiel des formats.

<sup>7</sup> Pour plus d'informations sur le processus d'import du référentiel, consulter le document *Modèle de workflow*, chapitre 5.III « Workflow d'administration d'un référentiel des formats ».



	Mise à jour d'un référentiel dont la date de création est antérieure à la date de création du référentiel préalablement importé.
Échec	Sans journalisation : - ré-import d'un référentiel dont le format ne correspond pas au format XML.
	Avec journalisation : - mise à jour d'un référentiel dans lequel un format ne déclare pas de PUID ou dont le champ PUID ne contient pas de valeur ; - mise à jour d'un référentiel dans lequel un format ne dispose pas d'intitulé ou dont le champ Name ne contient pas de valeur ; - mise à jour d'un référentiel dans lequel deux formats déclarent le même PUID, alors que ce dernier doit être unique.

### **Point d'attention :**

- Seules les opérations en succès et en avertissement sont journalisées dans le journal des opérations ;
- dans les cas d'un import ou ré-import en échec, l'action n'est pas journalisée.

Un rapport au format JSON est généré et précise les éléments suivants :

- le type d'opération, dont la valeur est « STP\_REFERENTIAL\_FORMAT\_IMPORT » ;
- la date et l'heure du début de la génération du rapport ;
- l'identifiant de l'opération d'import à l'origine de ce rapport ;
- le statut de l'opération, dont la valeur est soit « OK », soit « WARNING », soit « KO » ;
- la référence à l'ancienne et à la nouvelle version du référentiel ;
- la date de l'ancienne et de la nouvelle version du référentiel ;
- la liste des formats dont les informations ont été modifiées, avec le différentiel entre les anciens et les nouveaux éléments ;
- la liste des PUID des formats ajoutés au référentiel ;
- la liste des PUID supprimés ;
- le motif d'un avertissement.

### **2.1.4. Conseils de mise en œuvre**

À l'issue de cette phase de réalisation de fonctionnalités concernant le référentiel des formats, l'équipe projet Vitam est en mesure de fournir plusieurs recommandations de mise en œuvre.

#### **2.1.4.1. À quoi sert le référentiel des formats ?**

L'objectif du référentiel est de :

- connaître l'ensemble des formats pouvant être gérés et conservés par la solution logicielle Vitam ;
- permettre l'utilisation de l'outil d'identification de formats retenu à ce jour, Siegfried, et de pallier les défauts de cet outil, à savoir d'être en mesure, via des requêtes API, de :

- lister les formats connus de son référentiel interne et qu'il est donc en mesure de pouvoir identifier ;
- dire si un PUID est connu de son référentiel interne.

Il a également pour vocation d'être utilisé par la solution logicielle Vitam pour :

- effectuer des contrôles sur les formats des objets numériques associés aux archives prises en charge par le système. Dans le cas présent, le contrôle est fait au moyen de l'outil d'identification Siegfried et du référentiel des formats qu'il intègre et non pas à partir du référentiel interne de la solution logicielle Vitam ;
- sélectionner les formats dans :
  - les contrats d'entrées ;
  - les scénarios de préservation.

#### 2.1.4.2. Quand et comment créer un référentiel des formats ?

La création d'un référentiel des formats est un préalable au transfert d'objets numériques dans la solution logicielle Vitam. C'est pourquoi, lors de l'installation de la solution logicielle Vitam, un référentiel est initialisé par défaut. Sa version est compatible avec celle du référentiel importés dans l'outil d'identification de formats (Siegfried à ce jour). **Un administrateur fonctionnel n'a pas besoin, a priori, d'importer un référentiel des formats.** Il s'agit d'un acte d'exploitation technique.

#### **Points d'attention :**

- Après initialisation de la plate-forme, il est recommandé à l'administrateur technique de vérifier que la version du référentiel correspond à celle qui a été chargée dans l'outil d'identification Siegfried<sup>8</sup>.
- Au terme de la release 9, la solution logicielle Vitam supporte les **versions 91 et suivantes** du registre PRONOM.

#### 2.1.4.3. Quand et comment ré-importer et mettre à jour le référentiel des formats ?

The National Archives (UK) produit mensuellement une nouvelle version du fichier de signatures. Cette nouvelle version est alors publiée sur le site de l'institution, sauf dans deux cas :

- aucune signature n'a été ajoutée dans le fichier de signatures ;
- l'institution est en train de réaliser une mise à jour du fichier et préfère repousser sa publication.

Il est possible de mettre à jour régulièrement le référentiel dans la solution logicielle Vitam, en procédant à un réimport complet du référentiel, passant par l'import de la version du fichier de signatures nouvellement publiée. **Cette opération s'effectue uniquement sur le tenant d'administration.**

La mise à jour du référentiel s'avère nécessaire pour répondre aux besoins suivants :

- une montée de version de l'outil d'identification de formats (actuellement Siegfried) rend nécessaire la mise à jour d'une version plus récente du fichier de signature ;

---

<sup>8</sup> Pour plus d'informations, consulter *Document d'installation*, chapitre 6.1.4 « Post-installation : administration fonctionnelle ».

- on souhaite ajouter un format jusque-là absent du référentiel des formats, en l’ajoutant préalablement au registre PRONOM.
  - Soit ce nouveau format est intégré dans une version plus récente du registre PRONOM ;
  - Soit il a été ajouté spécifiquement dans un fichier de signatures.

De fait, il n’est pas obligatoire de monter de version du fichier de signatures dès qu’une nouvelle version paraît.

**Points d’attention :**

- Après la mise à jour du référentiel, l’administrateur technique doit vérifier que la version du référentiel correspond bien à celle qui a été chargée dans l’outil d’identification de formats Siegfried ou de procéder à sa mise à jour<sup>9</sup>.
- La solution logicielle Vitam permet d’ajouter des formats non référencés dans le référentiel PRONOM.
  - Dans ce cas précis, il faudra s’assurer que ces formats soient inclus dans la version du référentiel chargée dans Siegfried.
  - Par ailleurs, dans le cas d’une nouvelle version du registre PRONOM à installer dans la solution logicielle Vitam, il ne faudra pas oublier de maintenir les ajouts ou modifications spécifiques dans le nouveau fichier. En d’autres termes, lors d’une nouvelle version du registre PRONOM à importer dans le système, il faudra :
    - vérifier que cette nouvelle version n’intègre pas ces formats jusque-là non référencés ;
    - Si ce n’est pas le cas, rajouter ces formats spécifiques à PRONOM, afin qu’ils continuent à être gérés par le système.
- Si un format est absent ou a été ajouté dans le registre, ou si des informations sont incomplètes, il est nécessaire de rentrer en contact avec The National Archives (UK) et de proposer des mises à jour, afin que celles-ci soient intégrées à une version ultérieure du fichier de signatures<sup>10</sup>.

2.1.4.4. Comment procéder lors d’une mise à jour du référentiel des formats ?

Pour mettre à jour le référentiel des formats, il est recommandé de suivre les étapes suivantes :

Qui ?	Quoi ?	Via l’IHM dém Vitam ?
Administrateur fonctionnel	souhaite mettre à jour le référentiel des formats et propose à l’administrateur technique une version du référentiel à mettre à jour.	Non
Administrateur fonctionnel et/ou	Dans le cas de l’ajout de format(s) spécifique(s), non référencé(s) dans le registre PRONOM, ajoute le(s) format(s) concerné(s) dans le fichier	Non

<sup>9</sup> La procédure est expliquée dans *Documentation d’exploitation*, chapitre 5.10 « Montée de version du fichier de signature de Siegfried ».

<sup>10</sup> Les demandes peuvent être soumises à l’adresse suivante : <https://www.nationalarchives.gov.uk/contact-us/submit-information-for-pronom/> (Lien consulté le : 8 mars 2019).

technique	de signatures fourni par The National Archives (UK).	
Administrateur technique	vérifie si l’outil d’identification de formats Siegfried supporte la version du référentiel à mettre à jour ou si la mise à jour va nécessiter une montée de version de l’outil.	Non
Administrateur fonctionnel	recupère en ligne et importe le fichier de signatures sur le tenant d’administration, si l’outil d’identification de formats Siegfried supporte la version du référentiel à mettre à jour, l’administrateur fonctionnel.	Oui
Administrateur technique	met à jour le référentiel PRONOM dans l’outil d’identification de formats Siegfried. Si l’outil ne supporte pas la nouvelle version du référentiel, il monte de version afin que l’outil d’identification de formats Siegfried puisse supporter cette nouvelle version du référentiel.	Non

#### 2.1.4.5. Comment utiliser le référentiel des formats ?

Intitulé	Description	Niveau de recommandation
Import ou modification du référentiel des formats	Il est obligatoire d’effectuer les opérations d’import et de mise à jour du référentiel des formats sur le tenant d’administration.	Obligatoire
Import initial du référentiel des formats	Il est obligatoire d’importer le référentiel des formats lors de l’installation de la solution logicielle Vitam. La version de ce référentiel doit être identique à celle qui est importée dans l’outil Siegfried.	Obligatoire
Modification du fichier de signatures	Il est possible d’ajouter des formats non référencés dans le registre PRONOM. Dans ce cas précis, il faudra s’assurer que ces formats soient inclus dans la version du référentiel chargée dans Siegfried. Par ailleurs, dans le cas d’une nouvelle version du registre PRONOM à installer dans la solution logicielle Vitam, il ne faudra pas oublier de maintenir les ajouts ou modifications spécifiques qui ont été précédemment ajoutés dans le nouveau fichier.	Possible
Modification du fichier de signatures	Il n’est pas recommandé de modifier ou corriger les contenus présents dans le fichier de signatures fourni par The National Archives (UK), que l’on souhaite importer dans la solution logicielle Vitam.	Non recommandé
Modification du fichier de signatures	Si un format est absent du référentiel ou si des informations sont incomplètes, il est nécessaire de rentrer en contact avec The National Archives (UK) et de proposer des mises à jour, afin que celles-ci soient intégrées à une version ultérieure du fichier de signatures.	Recommandé

Mise à jour du référentiel des formats	Il est possible de mettre à jour le référentiel des formats. La version de ce nouveau référentiel doit être identique à celle qui est importée dans l’outil Siegfried. Si ce n’est pas le cas, il faut également mettre à jour le fichier de signatures importé dans cet outil.	Obligatoire
Mise à jour du référentiel des formats	Il est fortement recommandé de mettre à jour le référentiel des formats en le remplaçant par une version plus récente et non pas par une version plus ancienne. En effet, lors de cette mise à jour, la solution logicielle Vitam ne contrôle pas si les formats ôtés du référentiel correspondent à des formats utilisés par des objets qu’elle conserve.	Recommandé

## 2.2. Outils de préservation (griffons)

### 2.2.1. Définition

Les outils de préservation, appelés griffons au sein de la solution logicielle Vitam, sont des outils logiciels permettant d’effectuer des traitements sur les objets binaires en fonction de leur format.

Ces traitements peuvent être de différentes natures :

- analyse (ex : validation de formats, analyse de la qualité d’une image) ;
- génération d’un nouvel objet binaire (ex : conversion de format, création d’un autre usage, création d’un nouvel objet sans certaines métadonnées) dans un usage défini (BinaryMaster, Dissemination, Thumbnail, TextContent) ;
- extraction de métadonnées (métadonnées techniques, métadonnées descriptives) ;
- identification de format.

Un même griffon peut être utilisé pour effectuer plusieurs de ces traitements.

Plusieurs griffons peuvent être utilisés pour effectuer un traitement identique sur un même périmètre.

La solution logicielle Vitam implémente nativement plusieurs griffons :

- ImageMagick (*outil implémenté dans la release 9*) ;
- JHOVE (*outil implémenté dans la release 9*) ;
- Siegfried (*outil implémenté dans la release 9*) ;
- LibreOffice (*outil implémenté dans la release 9*) ;
- Tesseract (*outil implémenté dans la release 9*) ;
- ODFValidator (*outil implémenté dans la release 9*) ;
- VeraPDF (*outil implémenté dans la release 9*) ;
- OpenXMLValidator (*outil non implémenté à ce jour*) ;
- BFFValidator (*outil non implémenté à ce jour*).

## 2.2.2. Formalisation

### 2.2.2.1. Dans un fichier JSON

Un référentiel des griffons liste l'ensemble des griffons pouvant être utilisés par la solution logicielle Vitam. Ce référentiel est importé sous la forme d'un fichier JSON, contenant systématiquement l'ensemble des griffons utilisés.

```
{
  "Identifiant": "GRI-000001",
  "Name": "Griffon 1 : ImageMagick",
  "ExecutableName": "imagemagick-griffin",
  "ExecutableVersion": "V1.0.0"
}
```

Un griffon donné doit nécessairement être décrit avec les informations suivantes :

- identifiant signifiant du griffon, généré par l'application à l'origine de sa création (Identifiant) ;
- nom du griffon (Name) ;
- nom technique du griffon, utilisé pour lancer son exécution dans la solution logicielle Vitam (ExecutableName) ;
- version du griffon (ExecutableVersion) ;

Une description (Description) et une date de création (CreationDate), facultatives, peuvent venir compléter ces informations.

### 2.2.2.2. Dans la solution logicielle Vitam

Les griffons sont enregistrés dans la base de données MongoDB, dans la collection « Griffin », sous la forme d'enregistrements au format JSON.

Chaque enregistrement est modélisé comme suit<sup>11</sup> :

Champ	Description
_id	identifiant unique dans l'ensemble du système, fourni par ce dernier (obligatoire)
Identifiant	identifiant signifiant du griffon, généré par l'application à l'origine de sa création (champ obligatoire)
Name	nom du griffon (champ obligatoire)
Description	description du griffon (champ facultatif)
CreationDate	date de création du griffon. Si l'information n'est pas présente dans le fichier JSON importé dans la solution logicielle Vitam, la date est fournie par la solution logicielle Vitam et correspond à la date de création de l'enregistrement (champ obligatoire)
LastUpdate	date correspondant à la dernière modification de l'enregistrement, fournie par la solution logicielle Vitam (champ obligatoire)

<sup>11</sup> Pour plus d'informations, consulter *Modèle de données*, chapitre 10, « Collection Griffin ».

ExecutableName	nom technique du griffon, utilisé pour lancer son exécution dans la solution logicielle Vitam (champ obligatoire)
ExecutableVersion	version du griffon (champ obligatoire)
_tenant	tenant dans lequel l'enregistrement a été créé, fourni par le système (champ obligatoire)
_v	version du griffon, fournie par le système (champ obligatoire). <ul style="list-style-type: none"> <li>• Si l'enregistrement est égal à « 0 », il s'agit de l'enregistrement d'origine.</li> <li>• Si l'enregistrement est supérieur à « 0 », l'enregistrement a fait l'objet de 1 à n mise(s) à jour.</li> </ul>

## 2.2.3. Mécanismes mis en œuvre dans la solution logicielle Vitam

### 2.2.3.1. Implémentation des griffons

Un griffon est construit autour d'un outil logiciel interfacé avec la solution logicielle Vitam. Au moyen d'une interface fichier simple, ce griffon prend un lot d'objets binaires de formats donnés à traiter ainsi que les arguments spécifiant comment les traiter et fournit un résultat au regard de ces éléments.

Un griffon peut effectuer plusieurs actions de préservation, en fonction de ses capacités et du paramétrage choisi. Les griffons peuvent être programmés facilement, sans connaissance approfondie de la solution logicielle Vitam, dans n'importe quel langage informatique.

À ce stade, les griffons développés par l'équipe Vitam sont tous dans un environnement Linux<sup>12</sup>, à l'exception de l'un d'entre eux dont l'environnement Windows a été émulé sous Linux<sup>13</sup>.

Griffon	Langage de programmation	Environnement	Commentaires
ImageMagick		Linux	
JHOVE	Librairie JAVA	Linux	
VeraPDF	Librairie JAVA	Linux	
LibreOffice		Linux	
Siegfried		Linux	
Tesseract	C++	Linux	
ODFValidator	Librairie JAVA patchée	Linux	Exécuté en Mono
OpenXMLValidator	Librairie C sharp	Linux	Exécuté en Mono sous Linux
BFFValidator		Windows	Exécuté en WineHQ sous Linux

<sup>12</sup> Au terme de la release 9, les seuls griffons disponibles sont ImageMagick et JHOVE.

<sup>13</sup> Une réflexion est en cours pour gérer sur des Workers Windows des griffons conçus pour fonctionner sur un environnement Windows.

Ces griffons constituent des éléments de défiance en termes de sécurité informatique et des mécanismes doivent être mis en place pour les surveiller, par exemple en cas de délais de réponse trop important ou de génération d'objets numériques de taille supérieure à la taille attendue. Une réflexion est en cours sur l'utilisation de la technologie conteneur (type Docker) afin de les sécuriser (*service non encore implémenté*).

L'installation et l'exploitation des griffons constituent un **acte d'administration technique**.

### 2.2.3.2. Gestion du référentiel des griffons

L'ensemble des griffons implémentés et supportés par la solution logicielle Vitam doivent être référencés dans le référentiel des griffons.

Ce référentiel est multi-tenant. Il est administrable et journalisé depuis le tenant d'administration.

#### *2.2.3.2.1. Import*

Dans la solution logicielle Vitam, il est possible d'importer un référentiel des griffons, fonctionnant en mode « esclave ».

Il s'agit d'une opération d'administration (« MASTERDATA »), tracée dans le journal des opérations de la solution logicielle Vitam sur le tenant d'administration<sup>14</sup>.

#### *2.2.3.2.2. Ré-import et mise à jour*

La modification du référentiel des griffons s'effectue par un réimport du fichier au format JSON, en mode « annule et remplace ». De fait, la solution logicielle Vitam permet de :

- ajouter un nouveau griffon ;
- modifier les informations associées à un griffon ;
- supprimer un griffon.

Cette action provoque la création d'une nouvelle version du référentiel. Les différentes versions du référentiel font l'objet d'une sauvegarde sur les offres de stockage utilisées par la solution logicielle Vitam.

Elle fait l'objet d'une journalisation dans le journal des opérations (opération d'administration de type « MASTERDATA »)<sup>15</sup>.

Lors de ce ré-import ou de cette mise à jour, l'opération peut aboutir aux statuts suivants :

<b>Statut</b>	<b>Motifs</b>
Succès	Opération réalisée sans rencontrer de problèmes particuliers.

<sup>14</sup> Pour plus d'informations sur le processus d'import du référentiel, consulter le document *Modèle de workflow*, chapitre 5.XII « Workflow d'administration d'un référentiel des griffons ».

<sup>15</sup> Pour plus d'informations sur le processus d'import du référentiel, consulter le document *Modèle de workflow*, chapitre 5.XII « Workflow d'administration d'un référentiel des griffons ».



Avertissement	Mise à jour d'un référentiel dont au moins un griffon est déclaré dans un scénario de préservation.
Échec <sup>16</sup>	Sans journalisation : - ré-import d'un référentiel sous la forme d'un fichier qui n'est pas au format JSON ; - import d'un référentiel dont au moins un des champs contient une injection HTML.
	Avec journalisation : - mise à jour d'un référentiel dans lequel deux griffons portent le même identifiant métier ; - mise à jour d'un référentiel dans lequel un griffon ne déclare pas d'identifiant, d'intitulé, de nom technique d'outil ou de version d'exécutable ; - import d'un référentiel dans lequel un champ ne contient pas de valeur. Il peut s'agir des champs : Identifier, Name, ExecutableName, ExecutableVersion ; - import d'un référentiel dans lequel une valeur ne correspond pas au type d'indexation du champ défini dans l'ontologie (ex : valeur textuelle dans un champ de type « DATE »).

## 2.2.4. Conseils de mise en œuvre

À l'issue de cette phase de réalisation de fonctionnalités concernant le référentiel des griffons, l'équipe projet Vitam est en mesure de fournir quelques recommandations de mise en œuvre.

### 2.2.4.1. À quoi sert le référentiel des griffons ?

L'objectif du référentiel est de :

- connaître l'ensemble des griffons installés, gérés et utilisables dans la solution logicielle Vitam ;
- permettre l'utilisation des griffons ainsi référencés dans un à plusieurs scénario(s) de préservation.

### 2.2.4.2. Quand et comment créer un référentiel des griffons ?

#### 2.2.4.2.1. Au moment de l'installation de la solution logicielle Vitam

La solution logicielle Vitam permet, au moyen d'un paramétrage technique, d'installer un certain nombre de griffons, au moment de l'installation de la plate-forme<sup>17</sup>. Elle met, en effet, à disposition les griffons suivants :

- ImageMagick (*outil implémenté dans la release 9*) ;
- JHOVE (*outil implémenté dans la release 9*) ;
- Siegfried (*outil implémenté dans la release 9*) ;
- LibreOffice (*outil implémenté dans la release 9*) ;
- Tesseract (*outil implémenté dans la release 9*) ;

<sup>16</sup> Des précisions sur les messages d'erreur sont apportées dans l'annexe 2 « Messages d'erreur » du présent document.

<sup>17</sup> Pour plus d'informations, se référer à *Documentation d'installation*, chapitre 4.2.4.2 « Installation des greffons ».

- ODFValidator (outil implémenté dans la release 9) ;
- VeraPDF (outil implémenté dans la release 9) ;
- OpenXMLValidator (outil non implémenté à ce jour) ;
- BFFValidator (outil non implémenté à ce jour).

Leur installation est facultative. En d’autres termes, si on ne souhaite pas utiliser de griffons, pour des raisons sécuritaires ou fonctionnelles, il suffit de ne pas les installer.

Il est également possible de :

- ne pas installer la totalité des griffons mis à disposition ;
- installer des griffons supplémentaires<sup>18</sup>.

Cette opération relève d’un **acte technique**. Elle doit être suivie de la déclaration du griffon installé dans le référentiel des griffons, afin que le griffon puisse être utilisé lors d’une opération de préservation.

De fait, au moment de l’initialisation de la plate-forme, il est recommandé de :

Qui ?	Quoi ?	Via l’IHM dém Vitam ?
Administrateur fonctionnel	- définit ses besoins en préservation ; - le cas échéant, liste les griffons dont il a besoin.	Non
Administrateur technique	- vérifie si les griffons sont installés et disponibles par défaut dans la solution logicielle Vitam ; - le cas échéant, installe le(s) griffon(s) au moment de l’initialisation de la plate-forme.	Non
Administrateur technique et/ou fonctionnel	- crée le référentiel des griffons en y intégrant les griffons installés ; - importe le référentiel des griffons sur le tenant d’administration.	Oui

#### 2.2.4.2.2. Après installation de la solution logicielle Vitam

Tout paramétrage d’une opération de préservation passant par l’administration des scénarios de préservation nécessite l’installation préalable d’au moins un griffon dans la solution logicielle Vitam et sa déclaration dans le référentiel des griffons.

De fait, avant toute opération de préservation, il est recommandé de :

Qui ?	Quoi ?	Via l’IHM dém Vitam ?

<sup>18</sup> L’installation de griffons supplémentaires intervient après installation de la plate-forme. Sa mise en œuvre est expliquée dans *Documentation d’exploitation*, chapitre 5.11 « Griffins ».

Administrateur technique	- vérifie si le griffon que l'on souhaite solliciter est installé et disponible dans la solution logicielle Vitam ; - le cas échéant, installe le griffon <sup>19</sup> .	Non
Administrateur technique et/ou fonctionnel	- vérifie si le griffon est référencé dans le référentiel des griffons ; - le cas échéant, ajoute le griffon dans le référentiel des griffons	Oui
Administrateur technique et fonctionnel	crée un scénario de préservation utilisant le griffon en question	Oui
Administrateur fonctionnel	lance l'opération de préservation	Oui

### 2.2.4.3. Comment nommer un exécutable ?

Chaque griffon définit un nom technique faisant référence à l'outil exécutable par la solution logicielle Vitam (champ « ExecutableName »). Afin que l'outil soit exécuté :

- **si on utilise un griffon implémenté nativement dans la solution logicielle Vitam**, la règle est de nommer l'outil comme suit : nom de l'outil en minuscules et sans espace, suivi d'un tiret et du terme « griffin » ;

imagemagick-griffin jhove-griffin
--------------------------------------

- **si on utilise un griffon implémenté a posteriori dans la solution logicielle Vitam**, la règle est de nommer l'outil tel qu'il a été déclaré et installé dans la solution logicielle Vitam.

### 2.2.4.4. Quand et comment ré-importer et mettre à jour le référentiel des griffons ?

La mise à jour du référentiel des griffons s'avère nécessaire pour répondre aux besoins suivants :

- mise à jour des informations relatives à un ou plusieurs griffons (ex : modification de la description) ;
- ajout d'un nouveau griffon, pour faire suite à l'installation de ce dernier ;
- suppression d'un griffon.

La solution logicielle Vitam permet de mettre à jour le référentiel, en procédant à un réimport complet de celui-ci. **Cette opération s'effectue uniquement sur le tenant d'administration.**

#### **Points d'attention :**

- La montée de version d'un griffon peut engendrer la création d'un nouvel enregistrement dans le référentiel des griffons, quand l'outil a complètement changé et nécessite une nouvelle installation du griffon. En revanche, si sa montée de version ne consiste qu'en une mise à jour de certains de ses modules, il n'est pas nécessaire de modifier le référentiel des griffons.

<sup>19</sup> L'installation de griffons supplémentaires intervient après installation de la plate-forme. Sa mise en œuvre est expliquée dans *Documentation d'exploitation*, chapitre 5.11 « Griffins ».

- Après la mise à jour du référentiel, il est recommandé à l'administrateur technique de vérifier que la liste des griffons présents dans le référentiel correspond aux différents griffons installés sur la plate-forme.
- Un griffon pouvant être utilisé dans un ou plusieurs scénario(s) de préservation, il est recommandé de vérifier son utilisation dans le référentiel des scénarios de préservation, avant de procéder à sa suppression dans le référentiel des griffons.

#### 2.2.4.5. Comment procéder lors d'une mise à jour du référentiel des griffons ?

Pour mettre à jour le référentiel des griffons, il est recommandé de suivre les étapes suivantes :

Qui ?	Quoi ?	Via l'IHM démo Vitam ?
Administrateur fonctionnel	- émet le souhait d'ajouter un griffon et demande à l'administrateur technique de l'installer dans la solution logicielle Vitam.	Non
Administrateur technique	- procède à l'installation du griffon ; - fournit le nom de l'exécutable (ExecutableName) et sa version (ExecutableVersion) à l'administrateur fonctionnel.	Non
Administrateur fonctionnel ou technique	- met à jour le référentiel des griffons sur le tenant d'administration, en tenant compte des informations fournies par l'administrateur technique.	Oui

#### 2.2.4.6. Quand et comment supprimer un griffon ?

La solution logicielle Vitam ne permet pas de désinstaller un griffon, au sens technique du terme. Si on souhaite ne plus utiliser un griffon, il faut l'ôter du référentiel des griffons. Ainsi, il ne peut plus être utilisé dans un scénario de préservation.

Les raisons pouvant amener à supprimer un griffon sont les suivantes :

- Le griffon n'est jamais utilisé. Par souci de clarté, on souhaite le supprimer du référentiel ;
- Le griffon est obsolète et a nécessité l'installation d'une version plus récente. Plutôt que de maintenir deux griffons de versions différentes dans le référentiel, on souhaite en supprimer un du référentiel ;
- Il y a eu erreur de saisie, qui a entraîné la création d'un griffon qui n'est pas installé et/ou supporté par la solution logicielle Vitam.

**Point d'attention :** quand on supprime un griffon, il est recommandé de vérifier s'il est utilisé par un ou plusieurs scénario(s) de préservation et de modifier au préalable ces derniers avant de procéder à la suppression du griffon.

#### 2.2.4.7. Comment utiliser le référentiel des griffons ?

Intitulé	Description	Niveau de
----------	-------------	-----------

		recommandation
Import ou modification du référentiel des griffons	Il est obligatoire d'effectuer les opérations d'import et de mise à jour du référentiel des griffons sur le tenant d'administration.	Obligatoire
Import initial du référentiel des griffons	Il est recommandé d'importer un référentiel des griffons lors de l'installation de la solution logicielle Vitam, si on a installé des griffons. Les griffons listés dans le référentiel doivent correspondre à ceux qui ont été installés.	Recommandé
Mise à jour du référentiel des griffons	Il est recommandé de mettre à jour le référentiel des griffons. Les griffons listés dans le référentiel doivent correspondre à ceux qui ont été installés sur la plate-forme.	Recommandé
Nommage de l'exécutable	Il est obligatoire de nommer l'outil comme suit dans le référentiel des griffons : nom de l'outil en minuscules et sans espace, suivi d'un tiret et du terme « griffin », afin qu'il puisse être exécuté	Obligatoire
Supprimer un griffon	Il est recommandé de supprimer un griffon en l'enlevant du référentiel des griffons. En revanche, il n'est pas possible de le désinstaller de la plate-forme <sup>20</sup> .	Recommandé
Supprimer un griffon utilisé par un à plusieurs scénarios de préservation	Il est recommandé de vérifier l'utilisation d'un griffon dans le référentiel des scénarios de préservation, avant de procéder à sa suppression dans le référentiel des griffons. S'il est utilisé, il faut d'abord supprimer ou mettre à jour les scénarios de préservation où il est référencé, avant de procéder à sa suppression.	Recommandé

## 2.3. Scénarios de préservation

### 2.3.1. Définition

Un scénario de préservation décrit, pour une liste de formats donnée et un griffon donné, les actions de préservation à lancer (génération de binaires, analyse, identification, extraction de métadonnées). Pour chaque griffon, il définit des paramètres (arguments et conditions d'appel), permettant ainsi d'affiner l'action de préservation et d'aller au-delà, par exemple d'une simple génération d'objet binaire. En effet, un scénario de préservation peut avoir pour finalité les actions suivantes :

Action de préservation	Exemples
ANALYSE	la validation d'objets au format PDF par rapport aux spécifications du format PDF, en utilisant, par exemple, les griffons JHOVE ou VeraPDF ;

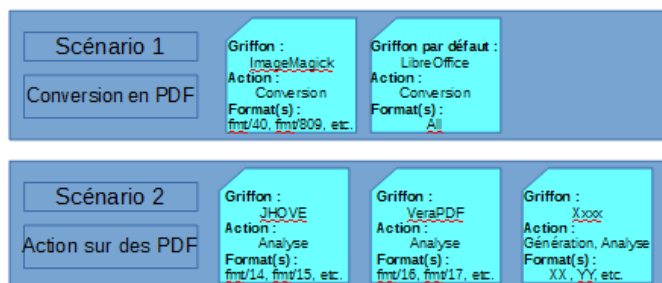
<sup>20</sup> Une réflexion est en cours afin de pouvoir désinstaller un griffon.

	la vérification qu'un objet a bien une taille de 1024x768 pixels.
GENERATE	la conversion d'originaux numériques au format bureautique vers un format PDF, en utilisant, en fonction des formats, soit ImageMagick avec des arguments propres à cet outil, soit LibreOffice avec des arguments propres à ce dernier outil
	la génération d'objets d'usage de diffusion à partir des originaux numériques en utilisant le griffon ImageMagick et avec pour cible la génération de fichiers JPEG ayant une taille maximale de 1024x768 pixels et une basse définition. Dans ce cas précis, l'objet généré n'est plus une réplique conforme de l'objet original.
IDENTIFY	la réidentification d'objets, en utilisant l'outil d'identification de formats Siegfried ( <i>fonctionnalité implémentée dans la release 9 – version 2.1.3</i> )
EXTRACT	l'extraction de métadonnées techniques en vue d'alimenter les métadonnées associées aux objets au moyen d'ImageMagick ( <i>fonctionnalité implémentée dans la release 9 – version 2.1.3</i> ).

Afin de pouvoir être utilisé, un scénario de préservation doit être associé à :

- 1 à n action(s) de préservation ;
- 1 à n griffon(s) ;
- 1 à n format(s).

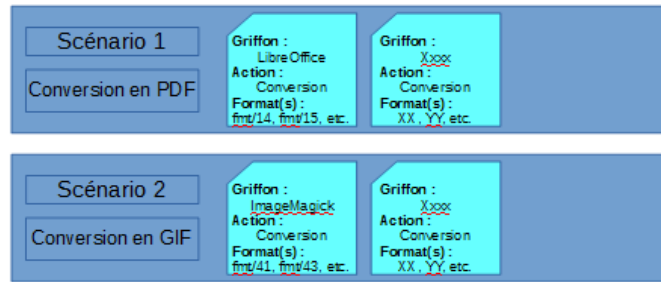
Exemple : le premier scénario couvre une seule action de préservation, effectuée par deux griffons différents sur des formats différents, tandis que le second porte sur deux actions différentes effectuée par des griffons différents sur des formats différents.



En outre, la solution logicielle Vitam permet d'utiliser :

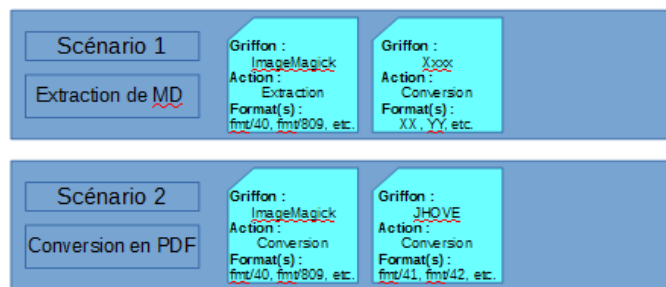
- une même action de préservation dans 1 à n scénario(s) de préservation (ex : plusieurs scénarios de préservation peuvent avoir pour objet une action de génération de binaires) ;

Exemple : deux scénarios de préservation effectuent une action identique de génération de binaires. Ils se différencient par les formats appelés et par les griffons : LibreOffice pour le premier et ImageMagick pour le second.



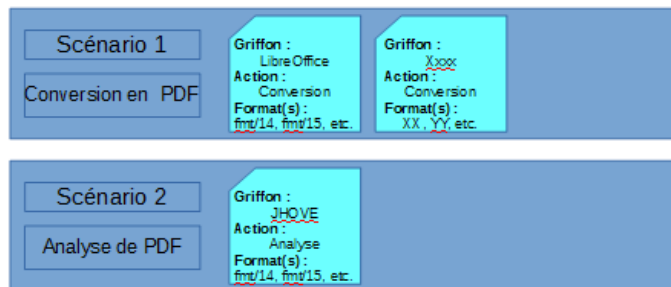
- un même griffon pour des tâches diverses dans 1 à n scénario(s) de préservation ;

*Exemple :* deux scénarios utilisent le griffon ImageMagick et listent les mêmes formats, mais l'un a pour action de l'extraction de métadonnées et l'autre de la génération de binaires.



- un même format dans 1 à n scénario(s) de préservation.

*Exemple :* deux scénarios de préservation ont pour objet les mêmes formats. En revanche, leurs actions et leurs griffons diffèrent.



## 2.3.2. Formalisation

### 2.3.2.1. Dans un fichier JSON

Un référentiel des scénarios de préservation liste l'ensemble des scénarios de préservation pouvant être utilisés par la solution logicielle Vitam. Ce référentiel est importé sous la forme d'un fichier JSON, contenant systématiquement l'ensemble des scénarios de préservation utilisés.

```
{
```

```
"Identifiant": "PSC-000001",
"Name": "Tranformation en GIF (fmt/43)",
"Description": "Ce scénario, appelant le griffon ImageMagick, permet de convertir des fichiers dont le PUID est
fmt/43 en fichiers au format GIF, et prenant la forme de vignettes de 100px x 100px",
"CreationDate": "2018-11-16T15:55:30.721",
"ActionList": [
  "GENERATE"
],
"GriffinByFormat": [
  {
    "FormatList": ["fmt/43"],
    "GriffinIdentifiant": "GRI-000001",
    "Timeout": 2000,
    "MaxSize": 10000000,
    "Debug": true,
    "ActionDetail": [
      {
        "Type": "GENERATE",
        "Values": {
          "Extension": "GIF",
          "Args": [
            "-thumbnail",
            "100x100"
          ]
        }
      }
    ]
  }
]
},
"DefaultGriffin": {
  "GriffinIdentifiant": "GRI-000002",
  "Timeout": 2000,
  "MaxSize": 10000000,
  "Debug": true,
  "ActionDetail": [
    {
      "Type": "GENERATE",
      "Values": {
        "Extension": "GIF",
        "Args": [
          "-thumbnail",
          "100x100"
        ]
      }
    }
  ]
}
}
```

Un scénario de préservation donné doit nécessairement comporter les informations suivantes :

- identifiant signifiant du scénario de préservation, généré par l'application à l'origine de sa création (Identifiant – obligatoire) ;
- nom du scénario de préservation (Name – obligatoire) ;
- action(s) couverte(s) par le scénario de préservation (ActionList – obligatoire). Ce champ



peut contenir 1 à 4 valeurs :

- « GENERATE » : le scénario de préservation couvre une opération de génération d'un nouvel objet (par exemple, conversion d'un format) ;
- « ANALYSE » : le scénario de préservation permet d'analyser un ou plusieurs format(s) de fichier (par exemple, validation d'un format) ;
- « IDENTIFY » : le scénario de préservation permet de réidentifier un ou plusieurs format(s) de fichier (*service implémenté dans la release 9 – version 2.3.0*) ;
- « EXTRACT » : le scénario de préservation couvre la fonctionnalité d'extraction de métadonnées contenues dans les objets binaires (*service implémenté dans la release 9 – version 2.1.3*) ;
- un bloc **GriffinByFormat** (Action(s) couverte(s) – obligatoire) précise l'(les) action(s) couverte(s) par le scénario de préservation, pour un griffon donné et une liste de format(s) donnée.

Une description (Description), ainsi qu'une action et un griffon à exécuter par défaut (bloc **DefaultGriffin**), facultatifs, peuvent venir compléter ces informations.

#### 2.2.2.2. Dans la solution logicielle Vitam

Les scénarios de préservation sont enregistrés dans la base de données MongoDB, dans la collection « PreservationScenario », sous la forme d'enregistrements au format JSON.

Chaque enregistrement est modélisé comme suit<sup>21</sup> :

- identifiant unique dans l'ensemble du système, fourni par ce dernier (\_id – obligatoire) ;
- identifiant signifiant du scénario de préservation, généré par l'application à l'origine de sa création (Identifieur – obligatoire) ;
- nom du scénario de préservation (Name – obligatoire) ;
- description du scénario de préservation (Description – facultatif) ;
- date de création du scénario de préservation. Si l'information n'est pas présente dans le fichier JSON importé dans la solution logicielle Vitam, la date est fournie par la solution logicielle Vitam et correspond à la date de création de l'enregistrement (CreationDate – obligatoire) ;
- date correspondant à la dernière modification de l'enregistrement, fournie par la solution logicielle Vitam (LastUpdate – obligatoire) ;
- action(s) couverte(s) par le scénario de préservation (ActionList – obligatoire). Ce champ peut contenir 1 à 4 valeurs :
  - « GENERATE » : le scénario de préservation couvre une opération de génération d'un nouvel objet (par exemple, conversion d'un format) ;
  - « ANALYSE » : le scénario de préservation permet d'analyser un ou plusieurs format(s) de fichier (par exemple, validation d'un format) ;

---

<sup>21</sup> Pour plus d'informations, consulter *Modèle de données*, chapitre 13 « Collection PreservationScenario ».

- « IDENTIFY » : le scénario de préservation permet de réidentifier un ou plusieurs format(s) de fichier (*service implémenté dans la release 9 – version 2.3.0*) ;
- « EXTRACT » : le scénario de préservation couvre la fonctionnalité d'extraction de métadonnées contenues dans les objets binaires (*service implémenté dans la release 9 – version 2.1.3*) ;
- tenant dans lequel l'enregistrement a été créé, fourni par le système (`_tenant` – obligatoire) ;
- version du scénario de préservation, fournie par le système (`_v` – obligatoire).
  - Si l'enregistrement est égal à « 0 », il s'agit de l'enregistrement d'origine.
  - Si l'enregistrement est supérieur à « 0 », l'enregistrement a fait l'objet de 1 à n mise(s) à jour.

Deux blocs supplémentaires, pouvant être vide, détaille le périmètre du scénario de préservation :

- un bloc **GriffinByFormat** (Action(s) couverte(s) – obligatoire) précise l'(les) action(s) couverte(s) par le scénario de préservation, pour un griffon donné et une liste de format(s) donnée ;
- un bloc **DefaultGriffin** (Action par défaut – facultatif) permet de définir une(des) action(s) par défaut, si l'action principale du scénario de préservation ne s'exécute pas.

Chacun de ces blocs comprend les éléments suivants :

- identifiant du griffon à exécuter (`GriffinIdentifier` – obligatoire) ;
- temps maximal de traitement (`Timeout` – obligatoire) ;
- taille maximale des objets (`MaxSize` – obligatoire) ;
- debug (`Debug` – obligatoire).
  - Si la valeur est égale à « true », les anomalies rencontrées<sup>22</sup> sont récupérées dans les logs de la solution logicielle Vitam ;
  - Si la valeur est égale à « false », les anomalies rencontrées ne sont pas récupérées dans les logs de la solution logicielle Vitam ;
- les commande(s) à exécuter (`ActionDetail` – obligatoire)<sup>23</sup>, qui se déclinent en :
  - Type d'action à exécuter (`Type` – obligatoire). Ce champ peut contenir 1 à 4 valeurs :
    - « GENERATE » : le scénario de préservation couvre une opération de génération d'un nouvel objet (par exemple, conversion d'un format) ;
    - « ANALYSE » : le scénario de préservation permet d'analyser un ou plusieurs format(s) de fichier (par exemple, validation d'un format) ;
    - « IDENTIFY » : le scénario de préservation permet de réidentifier un ou plusieurs format(s) de fichier (*service implémenté dans la release 9 – version 2.1.3*) ;
    - « EXTRACT » : le scénario de préservation couvre la fonctionnalité d'extraction de

---

<sup>22</sup> Les anomalies rencontrées proviennent du griffon. Elles émanent soit de l'outil en tant que tel, soit d'un mauvais paramétrage de l'utilisateur.

<sup>23</sup> Différents cas de figure concernant les commandes possibles à exécuter sont présentés dans l'annexe 3 « Paramétrages des scénarios de préservation » du présent document.

métadonnées contenues dans les objets binaires (*service implémenté dans la release 9 – version 2.1.3*) ;

- Valeur(s) (Values – facultatif), précisant les commandes passées par le scénario de préservation au griffon. Ce champ contient les éléments suivants :
  - des argument(s) (Args – facultatif), correspondant à une commande passée au griffon selon un langage et une grammaire comprise par ce dernier ;
  - une extension de sortie (Extension – facultatif), permettant de préciser, dans le cadre d’une opération de génération d’objet binaire, l’extension du format généré attendue.

Le bloc GriffinByFormat contient en plus un élément listant le(s) format(s) concerné(s) par le scénario de préservation (FormatList – obligatoire).

### 2.3.3. Mécanismes mis en œuvre dans la solution logicielle Vitam

#### 2.3.3.1. Import

Dans la solution logicielle Vitam, il est possible d’importer un référentiel des scénarios de préservation, fonctionnant en mode « esclave », tenant par tenant.

Il s’agit d’une opération d’administration (« MASTERDATA »), tracée dans le journal des opérations de la solution logicielle Vitam<sup>24</sup>.

#### 2.3.3.2. Ré-import et mise à jour

La modification du référentiel des scénarios de préservation s’effectue par un réimport du fichier JSON, en mode « annule et remplace ». De fait, la solution logicielle Vitam permet de :

- ajouter un nouveau scénario ;
- modifier les informations associées à un scénario ;
- supprimer un scénario.

Cette action provoque la création d’une nouvelle version du référentiel. Les différentes versions du référentiel font l’objet d’une sauvegarde sur les offres de stockage utilisées par la solution logicielle Vitam.

Elle fait l’objet d’une journalisation dans le journal des opérations (opération d’administration de type « MASTERDATA »)<sup>25</sup>.

Lors de ce ré-import ou de cette mise à jour, l’opération peut aboutir aux statuts suivants :

Statut	Motifs
Succès	l’opération a été réalisée sans rencontrer de problèmes particuliers.

<sup>24</sup> Pour plus d’informations sur le processus d’import du référentiel, consulter le document *Modèle de workflow*, chapitre 5.XIII « Workflow d’administration d’un référentiel des scénarios de préservation ».

<sup>25</sup> Pour plus d’informations sur le processus d’import du référentiel, consulter le document *Modèle de workflow*, chapitre 5.XIII « Workflow d’administration d’un référentiel des scénarios de préservation ».

Échec <sup>26</sup>	Sans journalisation : - import ou ré-import d'un référentiel dont le format ne correspond pas au format JSON ; - import d'un référentiel dont au moins un des champs contient une injection HTML.
	Avec journalisation : - mise à jour d'un référentiel dans lequel deux scénarios portent le même identifiant métier ; - mise à jour d'un référentiel dans lequel un scénario ne comporte pas au moins un champ obligatoire. Il peut s'agir, par exemple, des champs : Identifier, Name ; - import d'un référentiel dans lequel un champ ne contient pas de valeur. Il peut s'agir, par exemple, des champs : Identifier, Name ; - import d'un référentiel dans lequel une valeur ne correspond pas au type d'indexation du champ défini dans l'ontologie (ex : valeur textuelle dans un champ de type « DATE »).

### 2.3.4. Conseils de mise en œuvre

À l'issue de cette phase de réalisation de fonctionnalités concernant le référentiel des scénarios de préservation, l'équipe projet Vitam est en mesure de fournir quelques recommandations de mise en œuvre :

#### 2.3.4.1. À quoi sert le référentiel des scénarios de préservation ?

L'objectif du référentiel est de :

- connaître l'ensemble des scénarios de préservation disponibles et utilisables dans la solution logicielle Vitam ;
- gérer et administrer une politique de préservation dont les scénarios de préservation sont une déclinaison ;
- permettre l'utilisation des scénarios de préservation ainsi référencés dans une à plusieurs opération(s) de préservation.

#### 2.3.4.2. Quand et comment créer un référentiel des scénarios de préservation ?

Créer un scénario de préservation est un pré-requis obligatoire pour lancer n'importe quelle opération de préservation.

La création d'un scénario de préservation nécessite :

- l'installation préalable d'au moins un griffon dans la solution logicielle Vitam et sa déclaration dans le référentiel des griffons ;
- l'installation préalable d'un référentiel des formats contenant les formats destinés à être référencés dans un scénario de préservation.

De fait, afin de créer un scénario de préservation, il est recommandé de :

<b>Qui ?</b>	<b>Quoi ?</b>	<b>Via l'IHM</b>
--------------	---------------	------------------

<sup>26</sup> Des précisions sur les messages d'erreur sont apportées dans l'annexe 2 « Messages d'erreur » du présent document.

		démo Vitam ?
Administrateur fonctionnel	- détermine le scénario de préservation qu'il veut mettre en œuvre et l'outil qu'il souhaite utiliser.	Non
Administrateur technique	- vérifie si le griffon que l'on souhaite solliciter est installé et disponible dans la solution logicielle Vitam - le cas échéant, installe le griffon <sup>27</sup> .	Non
Administrateur technique et/ou fonctionnel	- vérifie si le griffon est référencé dans le référentiel des griffons ; - le cas échéant, ajoute le griffon dans le référentiel des griffons	Oui
Administrateur technique et fonctionnel	crée un scénario de préservation utilisant le griffon en question	Oui
Administrateur fonctionnel	lance l'opération de préservation	Oui

### 2.3.4.3. Quand et comment ré-importer et mettre à jour le référentiel des scénarios de préservation ?

La mise à jour du référentiel des scénarios de préservation s'avère nécessaire pour répondre aux besoins suivants :

- mise à jour des informations relatives à un ou plusieurs scénario(s) de préservation (ex : ajout d'un format dans un scénario, correction de la description, etc.) ;
- ajout d'un nouveau scénario de préservation ;
- suppression d'un scénario de préservation.

La solution logicielle Vitam permet de mettre à jour le référentiel dans la solution logicielle Vitam, en procédant à un **réimport complet** du référentiel.

**Points d'attention :** La solution logicielle Vitam effectue des contrôles sur l'existence des griffons et formats déclarés dans un scénario de préservation. En d'autres termes, il faut veiller à ce que les griffons et formats déclarés dans un scénario existent dans le référentiel des griffons pour les uns et dans le référentiel des formats pour les autres, avant de procéder à la mise à jour du référentiel des scénarios de préservation, sans quoi cette opération échouera.

### 2.3.4.4. Quand et comment créer un scénario de préservation ?

#### *2.3.4.4.1. Réflexion autour des besoins en termes de préservation*

Avant de créer un scénario de préservation et de l'ajouter au référentiel des scénarios de préservation, il faut réfléchir aux actions de préservation que l'on souhaite mener :

<b>Identifier le contexte</b>	Quels sont les objets numériques sur lesquels il faut procéder à une opération de
-------------------------------	---

<sup>27</sup> L'installation de griffons supplémentaires intervient après installation de la plate-forme. Sa mise en œuvre est expliquée dans *Documentation d'exploitation*, chapitre 5.11 « Griffins ».

	préservation ? Pourquoi ? Existe-il un besoin légal ?
<b>Identifier le périmètre</b>	Y a-t-il une seule ou plusieurs séries d'objets numériques à traiter ? Quel est leur format ? Exemple : l'opération doit-elle couvrir : <ul style="list-style-type: none"> <li>• l'ensemble des objets, de formats multiples, issus de producteurs différents et conservés dans la solution logicielle Vitam ?</li> <li>• Un format de fichiers en particulier ?</li> </ul>
<b>Identifier l'action</b>	Quelle action souhaite-t-on réaliser ? Souhaite-t-on procéder à une succession d'actions ?
<b>Identifier l'outil (griffon)</b>	Quel outil utiliser ? S'il existe plusieurs outils pour une même action de préservation, souhaite-on en utiliser un ou plusieurs ?

En fonction du périmètre, de(s) l'action(s) à effectuer et de(s) l'outil(s) privilégié(s), la conception du scénario de préservation variera. Par exemple :

- On peut envisager de réaliser un scénario unique d'identification de formats en utilisant le griffon par défaut Siegfried. Ce scénario ne spécifie pas de formats en particulier dans la mesure où Siegfried couvre les formats référencés dans le registre PRONOM ;
- Si on veut faire une succession d'actions sur un fichier (d'abord le transformer de A en B, puis de B en C), il faudra faire deux scénarios, car on va intervenir sur 2 objets de format différent ;
- Si on veut faire deux actions sur le même fichier (analyse, puis conversion), on doit créer deux scénarios : l'un pour l'analyse, l'autre pour la conversion, car un même scénario ne peut avoir deux actions à réaliser portant sur le même format, à moins que la seconde action ne soit réalisée avec un griffon par défaut.

#### 2.3.4.4.2. Rédaction d'un scénario de préservation

La création d'un nouveau scénario de préservation s'effectue par un réimport complet du référentiel, auquel a été ajouté un scénario supplémentaire.

Elle obéit à des règles strictes :

- Le scénario de préservation doit avoir un identifiant signifiant (Identifier) **unique** dans le référentiel des scénarios de préservation. Cet identifiant est généré par l'application à l'origine de sa création ;
- Il doit comporter un nom (Name) ;
- Il doit préciser les **actions de préservation qu'il couvre** (ActionList).
  - Ces actions, qui sont au nombre de quatre, correspondent obligatoirement à la liste suivante : « GENERATE », « ANALYSE », « IDENTIFY », « EXTRACT » ;
  - Le scénario peut couvrir 1 à 4 de ces actions.
- Il précise les **actions couvertes pour un griffon donné et une liste de format(s) donnée** (GriffinByFormat). Il peut préciser une à plusieurs actions. Chacune de ces actions :
  - peut être couverte par un griffon différent ;

- peut être différente les unes des autres ;
- doit porter sur des formats différents. Cela signifie que chaque liste de formats associée à un griffon donné ne doivent pas contenir de formats identiques. Si c'est le cas, le format répété ne sera traité qu'une seule fois dans le cadre d'une opération de préservation exécutant le scénario de préservation.

Pour une action couverte par un griffon donné, il est obligatoire de paramétrer les éléments suivants :

- identifiant du griffon à exécuter (GriffinIdentifier). Le griffon doit avoir été préalablement installé dans la solution logicielle Vitam et référencé dans le référentiel des griffons, sans quoi le scénario de préservation ne fonctionnera pas ;
- temps maximal de traitement (Timeout). Il s'agit d'indiquer le temps en **secondes** au bout duquel la solution logicielle Vitam, en l'absence de réponse du griffon, arrêtera l'action de préservation. Le timeout est fonction du griffon et de la réactivité de ce dernier. Par exemple :
  - pour Siegfried, il est de l'ordre de la seconde ;
  - pour Tesseract, dans la mesure où l'on doit pouvoir traiter des objets au format TIFF multi-images assez volumineux, il est plutôt de l'ordre de 10 minutes ;
- taille maximale des objets (MaxSize) sur lesquels l'action de préservation peut être effectuée en utilisant ce scénario de préservation. Elle est exprimée en **octets**.
- debug (Debug).
  - Si la valeur est égale à « true », les anomalies rencontrées<sup>28</sup> sont récupérées dans les logs de la solution logicielle Vitam. Devant permettre d'analyser un dysfonctionnement dans le contexte de l'utilisation d'un griffon dans la solution logicielle Vitam, cette option ne doit être activée que dans un contexte de pré-production ;
  - Si la valeur est égale à « false », les anomalies rencontrées ne sont pas récupérées dans les logs de la solution logicielle Vitam.
- la(les) commande(s) à exécuter (ActionDetail), qui se décline(nt) en :
  - Type d'action à exécuter (Type – obligatoire). Ce champ doit contenir une de ces quatre valeurs :
    - « GENERATE » : le scénario de préservation couvre une opération de génération d'un nouvel objet (par exemple, conversion d'un format) ;
    - « ANALYSE » : le scénario de préservation permet d'analyser un ou plusieurs format(s) de fichier (par exemple, validation d'un format) ;
    - « IDENTIFY » : le scénario de préservation permet de réidentifier un ou plusieurs format(s) de fichier (*service implémenté dans la release 9 – version 2.3.0*) ;
    - « EXTRACT » : le scénario de préservation couvre la fonctionnalité d'extraction de métadonnées contenues dans les objets binaires (*service implémenté dans la release 9 – version 2.3.0*) ;
  - Valeur(s) (Values), précisant les commandes passées par le scénario de préservation au

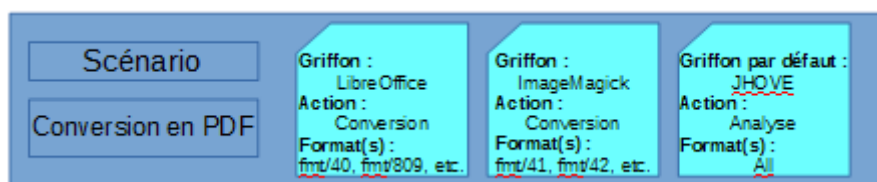
---

<sup>28</sup> Les anomalies rencontrées proviennent du griffon. Elles émanent soit de l'outil en tant que tel, soit d'un mauvais paramétrage de l'utilisateur.

griffon<sup>29</sup>. Ce champ contient des éléments différents en fonction du type d'action déclaré :

- pour les actions de préservation « ANALYSE », « GENERATE », « IDENTIFY » et « EXTRACT », ce champ a pour valeur « null » ou peut être absent ;
- pour l'action « GENERATE », c'est un objet pouvant, en plus du cas précédent, posséder deux champs :
  - une extension (Extension) chaîne de caractères servant à rajouter une extension aux fichiers générés, à écrire, de préférence, en minuscules (ex : pdf) ;
  - le cas échéant, une liste d'arguments (Args) utilisés lors de la commande système qu'effectue le griffon sur les objets concernés. Ces arguments doivent être conformes à ceux qui sont attendus et paramétrés pour être interprétés par le griffon utilisé<sup>30</sup>.

*Exemple : ce scénario de préservation permet de convertir au format PDF une liste de formats avec ImageMagick et une autre liste de formats avec LibreOffice. Pour les autres formats non listés, le scénario effectue par défaut une analyse avec JHOVE.*



```
[
  {
    "Identifiant": "PSC-000001",
    "Name": "Transformation en pdf",
    "Description": "Ce scénario permet de transformer un grand nombre de formats (bureautique et image) en PDF.",
    "ActionList": ["GENERATE"],
    "GriffinByFormat": [
      {
        "FormatList": ["fmt/40", "fmt/809"],
        "GriffinIdentifiant": "GRI-000003",
        "Timeout": 20,
        "MaxSize": 10000000,
        "Debug": true,
        "ActionDetail": [
          {"Type": "GENERATE", "Values": {"Extension": "pdf", "Args": ["-f", "pdf", "-e", "SelectedPdfVersion=1"]}},
          {"Type": "ANALYSE", "Values": {"Args": ["-strict"]}}
        ]
      },
      {
        "FormatList": ["fmt/41", "fmt/42"],
```

29 Les commandes possibles sont listées dans l'annexe 3 « Paramétrages des scénarios de préservation » du présent document.

30 Les commandes possibles sont listées dans l'annexe 3 « Paramétrages des scénarios de préservation » du présent document.



```
"GriffinIdentifieur": "GRI-000001",
"Timeout": 10,
"MaxSize": 10000000,
"Debug":true,
"ActionDetail": [
  {"type": "GENERATE","values": {"Extension": "pdf","Args": ["-quality","100","-units","PixelsPerInch","-
density","72x72"]}}}
]
},
"DefaultGriffin": {
  "GriffinIdentifieur": "GRI-000005",
  "Timeout": 10,
  "MaxSize": 10000000,
  "Debug":true,
  "ActionDetail": [
    {"Type": "ANALYSE","Values": {"Args": ["-strict"]}}}
  ]
}
}
```

#### 2.3.4.5. Quand et comment supprimer un scénario de préservation ?

Si on ne souhaite plus utiliser un scénario de préservation, il faut l'ôter du référentiel des scénarios de préservation. Ainsi, il ne peut plus être utilisé dans une opération de préservation.

Les raisons pouvant amener à supprimer un scénario sont les suivantes :

- Le scénario n'est jamais utilisé ou n'est plus utilisé. Par souci de clarté, on souhaite le supprimer du référentiel ;
- Le référentiel contient deux scénarios identiques. Plutôt que de les maintenir séparément, on souhaite en supprimer un du référentiel ;
- Le scénario fait appel à un griffon qui n'est plus utilisé dans le système ;
- Il y a eu erreur de saisie dans le référentiel, qui a entraîné la création d'une deuxième version d'un scénario de préservation.

**Point d'attention :** avant de supprimer tout scénario de préservation, il est recommandé de vérifier qu'il n'est pas utilisé dans le cadre d'une opération de préservation en cours de réalisation.

#### 2.3.4.6. Comment élaborer et structurer un référentiel des scénarios de préservation ?

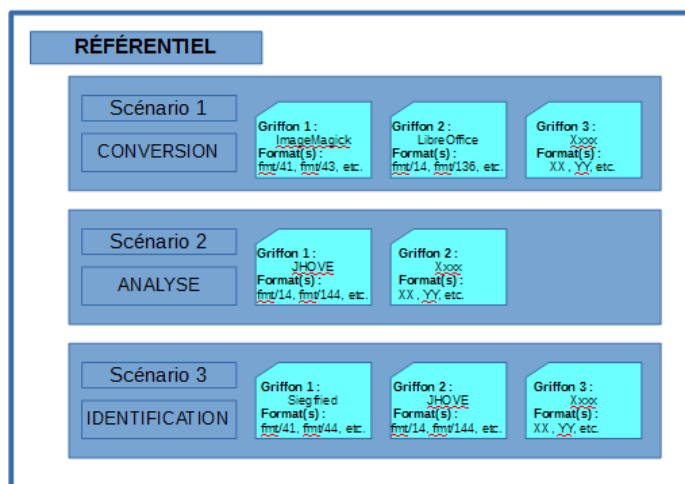
Le référentiel des scénarios de préservation peut être alimenté de manière très modulaire :

- d'une manière très générique, en réduisant le nombre de scénarios au nombre d'actions de préservation possibles : validation, conversion, identification, extraction ;
- d'une manière plus spécifique, en se focalisant sur :
  - un type de format en particulier par scénario ;
  - la finalité du scénario de préservation, en dissociant les scénarios par type d'action et format.

#### Cas 1 : un scénario par type d'action de préservation

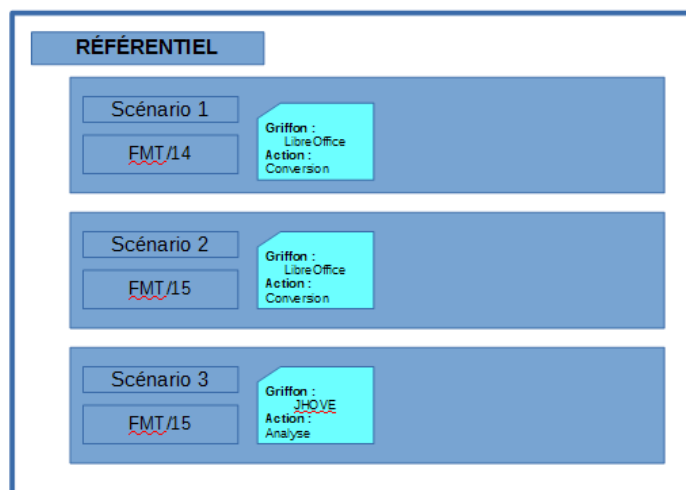
*Cette première solution présente l'avantage de limiter le nombre de scénarios. En revanche, la mise au point*

de chacun des scénarios peut s'avérer complexe et leur utilisation dans le cadre d'une opération de préservation peut s'avérer lourde en termes de traitement et de performance. L'élaboration de ce type de scénarios génériques peut être envisagée dans le cadre d'opérations de préservation portant sur l'ensemble des objets conservés dans la solution logicielle Vitam et tournant en tâche de fond.



### Cas 2 : un scénario par format

Cette deuxième solution permet d'effectuer des opérations de préservation plus ciblées et de répondre à des besoins plus circonscrits en termes de préservation. L'avantage de cette solution est de permettre de ne pas lancer une opération sur l'ensemble des objets conservés dans la solution logicielle Vitam et d'avoir une vision plus claire du contenu de chacun des scénarios de

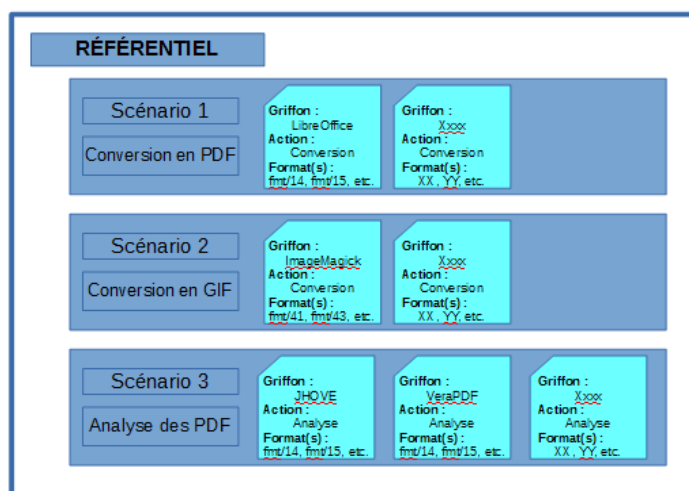


préservation.

### Cas 3 : un scénario par finalité

Cette troisième solution, modulaire, permet de disposer de scénarios génériques, portant sur l'ensemble des formats supportés dans la solution logicielle Vitam, en vue de réaliser une opération

de préservation en particulier (ex : générer des GIF, grâce à plusieurs griffons et sur une liste exhaustive de formats). Mais elle permet également d’effectuer des actions plus ciblées et de répondre à des besoins plus circonscrits (ex : valider des fichiers PDF).



### 2.3.4.7. Comment utiliser le référentiel des scénarios de préservation ?

Intitulé	Description	Niveau de recommandation
Import du référentiel des scénarios de préservation	<p>Si l’on souhaite procéder à des opérations de préservation, il est obligatoire de :</p> <ul style="list-style-type: none"> <li>• avoir préalablement installé au moins un griffon,</li> <li>• avoir importé un référentiel des griffons, déclarant au moins ce griffon,</li> <li>• avoir paramétré et importé au moins un scénario de préservation utilisant ce griffon.</li> </ul>	Obligatoire
Import et/ou mise à jour du référentiel des scénarios de préservation	<p>La solution logicielle Vitam effectue des contrôles sur l’existence des griffons et formats déclarés dans un scénario de préservation. En d’autres termes, il est obligatoire de veiller à ce que les griffons et formats déclarés dans un scénario existent dans le référentiel des griffons pour les uns et dans le référentiel des formats pour les autres, avant de procéder à la mise à jour du référentiel des scénarios de préservation, sans quoi cette opération échouera.</p>	Obligatoire
Activer le debug	<p>Le debug permet d’analyser un dysfonctionnement dans le contexte de l’utilisation d’un griffon dans la solution logicielle Vitam, que ce soit une anomalie émanant de l’outil ou d’un mauvais paramétrage de l’utilisateur.</p> <p>Il n’a pas vocation à être utilisé sur une plate-forme en production. En revanche, son usage est recommandé dans un cadre de préproduction.</p>	Recommandé

<p>Supprimer un scénario de préservation</p>	<p>La solution logicielle Vitam permet de supprimer un scénario en l'enlevant du référentiel des scénarios de préservation. Avant de supprimer tout scénario de préservation, il est recommandé de vérifier qu'il n'est pas utilisé dans le cadre d'une opération de préservation.</p>	<p>Recommandé</p>
--	--	-------------------

## 3. Entrées et préservation

En entrée dans le système, la solution logicielle Vitam permet de :

- paramétrer des **filtres au moyen des contrats d'entrée** afin de contraindre les objets pouvant y entrer ;
- effectuer des **contrôles génériques**, notamment d'identification, dans le cadre d'opérations d'entrée dans le système.

### 3.1. Définition

Assurer la préservation et garantir un archivage pérenne commence dès l'entrée des objets dans le système. Afin d'atteindre cet objectif, et ce en fonction du degré de contrôle souhaité, il peut être envisagé de :

- imposer ou non une liste de formats autorisés ;
- accepter ou non des formats non identifiés ;
- contrôler les objets transférés dans la solution logicielle Vitam en termes d'usages ;
- vérifier l'identification des objets, voire les réidentifier ;
- valider les formats de ces objets (*service non implémenté*).

Pour ce faire, dans le cadre du processus d'entrée d'un ensemble d'archives, la solution logicielle propose différentes fonctionnalités :

- elle effectue **automatiquement et de manière générique** des tâches d'identification ;
- elle permet, le cas échéant, de rajouter des vérifications supplémentaires au moyen d'un plug-in ;
- elle propose également des **paramétrages**, adaptables en fonction des besoins, permettant de contrôler les objets transférés, au moyen des contrats d'entrée.

### 3.2. Mécanismes mis en œuvre par la solution logicielle Vitam

#### 3.2.1. Filtres sur les entrées

La solution logicielle offre la possibilité de paramétrer des filtres à appliquer sur les bordereaux de transfert et les objets qu'ils référencent au moyen des contrats d'entrée.

En effet, un contrat d'entrée permet d'effectuer des contrôles sur les groupes d'objets transférés dans la solution logicielle Vitam au moyen de quatre options<sup>31</sup> :

- Une première option permet d'autoriser ou non le transfert de groupes d'objets ne contenant pas d'objets de type « Master » :
  - Si sa valeur est égale à « true », le bordereau de transfert devra obligatoirement contenir des objets de type « Master », qu'ils soient binaires (« BinaryMaster ») ou physiques (« PhysicalMaster ») ;

---

<sup>31</sup> Dans le cadre des prochains travaux de réalisation, il est prévu de pouvoir paramétrer et activer des scénarios de préservation au moyen des contrats d'entrée.

- Si sa valeur est égale à « false », le bordereau de transfert sera autorisé à contenir des groupes d'objets sans objet de type « Master » ;
- Si le contrat d'entrée, lors de son initialisation, ne détermine pas l'option retenue, la solution logicielle Vitam appliquera par défaut une valeur égale à « true » et imposera la présence d'objets de type « Master » dans les bordereaux de transfert.
- Une deuxième option permet de déterminer quel(s) type(s) ou usage(s) d'objets sont attendus dans les bordereaux de transfert (BinaryMaster, Dissemination, Thumbnail, TextContent), dans le cas où l'on souhaite rattacher un objet à un groupe d'objets déjà conservé dans la solution logicielle Vitam :
  - Si le contrat d'entrée permet le transfert de n'importe quel type d'usage, le bordereau de transfert pourra contenir n'importe quel type d'usage ;
  - Si le contrat d'entrée précise quel(s) usage(s) il autorise, le bordereau de transfert devra nécessairement contenir le(s) seul(s) usage(s) déclaré(s) dans le contrat d'entrée, sans quoi le transfert échouera ;
  - Si le contrat d'entrée, lors de son initialisation, ne détermine pas l'option retenue, la solution logicielle Vitam acceptera n'importe quel usage en entrée.
- Une troisième option permet de déterminer quel(s) format(s) de fichiers sont attendus dans les bordereaux de transfert :
  - Si le contrat d'entrée permet le transfert de n'importe quel format de fichier, le bordereau de transfert pourra contenir des objets de n'importe quel format possible ;
  - Si le contrat d'entrée précise quel(s) format(s) de fichier il autorise, le bordereau de transfert devra obligatoirement contenir des objets conformes au(x) seul(s) format(s) déclaré(s) dans le contrat d'entrée, sans quoi le transfert échouera ;
  - Si le contrat d'entrée, lors de son initialisation, ne détermine pas l'option retenue, la solution logicielle Vitam acceptera n'importe quel format en entrée.
- Une quatrième option permet d'autoriser le transfert d'objets dont le format n'est pas identifiable au moyen d'un paramétrage :
  - Si sa valeur est égale à « true », il sera possible de transférer des objets dont l'outil d'identification Siegfried est incapable d'identifier le format, malgré cette absence d'identification ;
  - Si sa valeur est égale à « false » (valeur par défaut), il ne sera possible de transférer dans la solution logicielle Vitam que des objets dont le format est identifié par l'outil d'identification Siegfried.

## 3.2.2. Contrôles sur les entrées

### 3.2.2.1. Service d'identification

Dans le cadre d'un transfert d'archives, la solution logicielle Vitam :

- demande à l'outil d'identification de formats Siegfried de procéder à l'identification des fichiers qui sont transférés dans un *Submission Information Package* (SIP) ;

- vérifie que l'identification retournée existe dans le référentiel des formats dont Siegfried dispose ;
- compare l'identification obtenue grâce à Siegfried avec l'identification proposée dans le manifeste du SIP ;
- peut :
  - rejeter les SIP dans lesquels des fichiers n'ont pas pu être identifiés (en retournant la ou les erreur(s) identifiée(s) au service demandeur dans le message ArchiveTransferReply) ;
  - corriger l'identification des fichiers, si l'identification obtenue grâce à l'outil d'identification de formats Siegfried diffère de celle proposée dans le bordereau de transfert ;
  - le cas échéant, accepter les entrées contenant des objets dont le format n'est pas identifié, si le contrat d'entrée l'autorise ;
- enregistre l'identification obtenue grâce à l'outil d'identification de formats Siegfried en conservant, dans le journal du cycle de vie des groupes d'objets correspondant, l'identification transmise dans le bordereau de transfert si celle-ci est différente.

Si le contrat d'entrée autorise le transfert d'objets au format inconnu, ceux-ci seront enregistrés avec une extension « .unknown ».

Ces contrôles permettent ainsi de :

- effectuer une identification la plus fine possible des formats de fichiers, en ne se contentant pas d'une identification sur la base de l'extension ;
- recourir aux services de l'outil d'identification de formats Siegfried. Ce choix a été effectué sur la base d'une analyse des avantages et des inconvénients de chaque outil, réalisée par le Service interministériel des Archives de France (SIAF) et l'équipe projet Vitam ;
- procéder à l'identification lors du processus de transfert et de prise en charge par la solution logicielle Vitam.

### 3.2.2.2. Service supplémentaire

En plus de ce processus d'identification générique, la solution logicielle offre la possibilité d'intégrer d'autres outils que celui qu'elle utilise nativement (Siegfried) via des plug-in<sup>32</sup>.

Les API mises en place permettent de dialoguer avec ces outils au moyen de 3 types de requêtes :

- récupérer l'identification du format d'un fichier envoyé à l'outil ;
- récupérer la liste des PUID connus du référentiel interne à l'outil ;
- vérifier si un format est connu du référentiel interne de l'outil sur la base de son PUID.

L'exécution de ces plug-in et son résultat peuvent être accompagnés par l'ajout d'étapes supplémentaires, **spécifiques et non optionnelles**, dans le workflow d'entrée.

#### **Point d'attention :**

- Il est recommandé de s'assurer que le référentiel interne au plug-in soit compatible avec le

---

<sup>32</sup> Dans le cadre des prochains travaux de réalisation, il est prévu de pouvoir paramétrer et activer des scénarios de préservation.

- référentiel des formats utilisé par la solution logicielle Vitam.
- L'installation de ce plug-in relève d'un acte technique.

### 3.3. Conseils de mise en œuvre

À l'issue de cette phase de réalisation de fonctionnalités, l'équipe projet Vitam est en mesure de fournir quelques recommandations de mise en œuvre :

#### 3.3.1. Quand et comment ajouter des contrôles sur les entrées ?

La solution logicielle Vitam permet d'ajouter des plug-in et de créer en parallèle de nouvelle(s) étape(s) dans le workflow, faisant état du traitement opéré par ce(s) plug-in(s)<sup>33</sup>. Le paramétrage de ces deux fonctionnalités relève d'un **acte technique**, réalisé :

- soit au moment de l'installation de la plate-forme,
- soit sur une plate-forme en activité. Cette opération peut nécessiter une interruption de service et un redémarrage de la plate-forme.

On peut envisager d'ajouter un plug-in et une étape au workflow d'entrée pour répondre aux besoins suivants :

- identifier les objets avec un outil différent de Siegfried ;
- valider automatiquement les objets en entrée du système.

#### **Point d'attention :**

- La nouvelle étape ainsi ajoutée s'insère dans le workflow d'entrée générique de la solution logicielle Vitam<sup>34</sup>.
  - Cette nouvelle étape n'est ni optionnelle, ni paramétrable au moyen des contrats d'entrée. De fait, elle devient générique.
  - Si on a installé un plug-in d'identification autre que Siegfried et une étape d'identification utilisant ce plug-in, dans le cadre du processus de transfert, la solution logicielle Vitam maintient néanmoins les étapes d'identification utilisant l'outil d'identification Siegfried. Dans ce cas-là :
    - le workflow d'entrée contiendra deux étapes d'identification, recourant aux services de deux plug-in différents ;
    - l'identification effectuée par le second plug-in n'est qu'informatrice et destinée à bloquer une entrée si l'outil échoue à reconnaître un format. En revanche, ce plug-in n'a pas vocation à se substituer à l'outil d'identification de formats utilisé par la plate-forme (Siegfried actuellement) en termes de réidentification en entrée.
- Il est recommandé de s'assurer que le référentiel interne au plug-in soit compatible avec le référentiel des formats utilisé par la solution logicielle Vitam.

---

<sup>33</sup> L'ajout d'un plug-in dans la solution logicielle Vitam, ainsi que la modification du workflow sont expliqués dans *Documentation d'exploitation*, chapitre 8.2.7 « Hello Word Plugin ».

<sup>34</sup> Dans le cadre des prochains travaux de réalisation, il est prévu de pouvoir paramétrer et activer des scénarios de préservation en fonction des besoins en termes de préservation et du niveau de service souhaité.



### 3.3.2. Comment paramétrer les filtres sur les entrées ?

Intitulé	Description	Niveau de recommandation
Application versant des originaux numériques	Pour une application devant transférer uniquement des originaux numériques, il est recommandé de contrôler au moyen du contrat d'entrée que les groupes d'objets qu'elle transfère contiennent obligatoirement un objet de type « Master » (« BinaryMaster »)	Recommandé
Application versant des objets d'un même usage à ajouter à des groupes d'objets déjà présents dans le système	Il est recommandé d'indiquer dans le contrat d'entrée quel est l'usage des objets à ajouter à des groupes d'objets déjà existant dans le système, si ces usages sont connus. Par exemple, un Portail Archives qui diffuse des copies numériques de diffusion pourrait ne devoir verser que des objets de type « Dissemination » pour compléter les groupes d'objets contenant les originaux.	Recommandé
Application versant des objets de différents usages à ajouter à des groupes d'objets déjà présents dans le système	Quand on ne connaît pas les usages des objets qui pourraient être transférés dans la solution logicielle pour compléter les groupes d'objets déjà présents dans le système ou quand on ne souhaite pas imposer un contrôle sur les usages de des objets à transférer, il est recommandé d'autoriser le transfert d'objets de tous les types d'usage.	Recommandé
Application versant des objets au(x) format(s) connu(s)	Il est recommandé d'indiquer dans le contrat d'entrée quel(s) est(sont) le(s) format(s) des archives à transférer par une application versante, si ce(s) format(s) sont connus ou si l'on souhaite imposer une liste de formats propre à l'archivage. Par exemple : <ul style="list-style-type: none"> <li>• un Portail Archives qui diffuse des copies numériques de diffusion pourrait ne devoir verser que des objets au format JPEG ;</li> <li>• un système d'archivage électronique ne souhaitant garantir la conservation d'objets qu'aux formats ODT, ODS, ODP, PDF, TIFF pourrait ne déclarer que des contrats d'entrée n'acceptant que ces seuls formats.</li> </ul>	Recommandé
Application versant des objets aux formats différents et non connus à l'avance	Quand on ne connaît pas les formats des archives qui pourraient être transférées dans la solution logicielle ou quand on ne souhaite pas imposer des formats pour les archives à transférer et quand on souhaite, néanmoins, que ces formats soient identifiés par la solution logicielle Vitam lors du transfert des archives, il est recommandé d'autoriser le transfert de groupes d'objets contenant tous les formats possibles.	Recommandé
Application versant des objets aux formats différents et non	Quand on ne connaît pas les formats des archives qui pourraient être transférées dans la solution logicielle ou quand on ne souhaite pas imposer des formats pour les archives à transférer	Recommandé

connus à l'avance	et quand on souhaite n'avoir aucun contrôle d'identification de format bloquant sur elles au moment de leur transfert dans la solution logicielle Vitam, il est recommandé d'autoriser le transfert de groupes d'objets contenant tous les formats possibles et de désactiver le contrôle d'identification des objets.	
-------------------	--	--

## 4. Gestion des archives et préservation

La solution logicielle Vitam offre plusieurs fonctionnalités permettant de gérer la pérennité des archives :

- des traitements sur les objets numériques au moyen d'une **opération générique de préservation** ;
- des audits vérifiant l'**existence** et l'**intégrité** des objets numériques, la **cohérence** des unités archivistiques et des objets numériques par rapport aux différents journaux produits dans la solution logicielle Vitam, et permettant leur **correction** en cas d'anomalie.

### 4.1. Préservation

#### 4.1.1. Définition

La solution logicielle Vitam intègre des fonctionnalités de préservation. Celles-ci consistent à appliquer un traitement sur un lot d'objets numériques définis par leur **usage** (BinaryMaster, Dissemination, Thumbnail, TextContent) et leur **numéro de version** (première version ou dernière version).

Ce traitement correspond à quatre types d'actions de préservation :

- **L'analyse** : cette action consiste à valider un format ou la qualité d'une image par rapport à des spécifications.  
Le résultat de cette action est une valeur contenue dans une énumération, exprimée dans un rapport. Par exemple :
  - Message dans le cas d'une validation de format : VALID, WELL\_FORMED, NON\_VALID, WRONG\_FORMAT...
  - Message dans le cas d'une analyse de la qualité d'une image : nette, floue...
- **la génération de binaire dans un usage défini** : il peut s'agir d'une conversion de format, de la création d'un autre usage ou de la création d'un nouvel objet sans certaines métadonnées.  
Le résultat de cette action est la création d'un nouvel objet binaire, ayant fait l'objet d'un calcul d'empreinte, d'une identification de format, d'un stockage et d'une indexation passant par la mise à jour de son groupe d'objets techniques, ainsi que d'une mise à jour du registre des fonds ;
- **l'extraction de métadonnées (service implémenté dans la release 9 – version 2.3.0)** : cette action consiste en la création de nouvelles métadonnées, issues des métadonnées techniques encapsulées dans les objets binaires.  
Ces nouvelles métadonnées sont enregistrées et indexées en tant que métadonnées de l'unité archivistique ou du groupe d'objets techniques<sup>35</sup> ;

---

<sup>35</sup> Dans le cadre de la release 9 (version 2.3.0), l'extraction de métadonnées alimente les métadonnées du groupe d'objets techniques.

- **L'identification de format (service implémenté dans la release 9 – version 2.3.0)** : il s'agit d'identifier ou de réidentifier a posteriori des fichiers déjà transférés dans la solution logicielle Vitam.

Exemples :

Action de préservation	Exemples
ANALYSE	la validation d'objets au format PDF par rapport aux spécifications du format PDF, en utilisant, par exemple, les griffons JHOVE ou VeraPDF.
	La vérification qu'un objet a bien une taille de 1024x768 pixels.
GENERATE	la conversion d'originaux numériques au format bureautique vers un format PDF, en utilisant, en fonction des formats, soit ImageMagick avec des arguments propres à cet outil, soit LibreOffice avec des arguments propres à ce dernier outil
	la transformation d'originaux numériques en objets d'usage de diffusion en utilisant le griffon ImageMagick et avec pour cible la génération de fichiers JPEG ayant une taille maximale de 1024x768 pixels et une basse définition. Dans ce cas précisément, l'objet généré n'est plus une réplique conforme de l'objet original.
IDENTIFY	la réidentification d'objets, en utilisant Siegfried ( <i>fonctionnalité implémentée dans la release 9 – version 2.3.0</i> )
EXTRACT	l'extraction de métadonnées techniques en vue d'alimenter les métadonnées associées aux objets au moyen d'ImageMagick ( <i>fonctionnalité implémentée dans la release 9 – version 2.3.0</i> ).

## 4.1.2. Mécanismes mis en œuvre dans la solution logicielle Vitam

### 4.1.2.1. Définition/Paramétrage de l'opération de préservation

La solution logicielle Vitam permet de réaliser une opération de préservation en associant à un lot d'archives plusieurs critères :

- des critères déterminés dans un **scénario de préservation** qui définit une ou plusieurs actions à exécuter, avec l'aide d'un griffon, sur un à plusieurs formats ;
- en déterminant sur quelle **version et usage** des objets numériques doit intervenir l'opération. Il peut s'agir de :
  - n'importe quel usage (original numérique, vignette, texte brut, version de diffusion) ;
  - la première version ou la dernière version de groupe(s) d'objet(s) ;
- un **usage cible** (original numérique, vignette, texte brut, version de diffusion).

### 4.1.2.2. Opération de préservation

Dans le cadre du processus de préservation d'un ensemble d'archives, suite à la réception d'une requête, la solution logicielle Vitam effectue les tâches et traitements de contrôles internes suivants<sup>36</sup> :

<sup>36</sup> Pour plus d'informations sur l'opération de préservation, consulter le document *Modèle de workflow*, chapitre 11.I « Workflow de

- **préparation** du traitement avec vérification des seuils de limitation des traitements portant sur les unités archivistiques et récupération des groupes d’objets techniques et du scénario de préservation ;
- **action** de préservation, elle-même subdivisée en différentes tâches :
  - lancement du griffon déclaré dans le scénario de préservation ;
  - dans le cas d’une génération de binaires :
    - identification du format des objets binaires générés au moyen de l’outil d’identification des formats Siegfried et vérification que le format identifié est bien référencé dans le référentiel des formats ;
    - calcul de l’empreinte des nouveaux objets binaires ;
    - écriture des objets générés sur les offres de stockage ;
    - indexation des métadonnées de préservation et mise à jour des journaux du cycle de vie concernés ;
    - sauvegarde des métadonnées et des journaux de cycle de vie sur les offres de stockage ;
  - dans le cas d’une réidentification ou d’une extraction de métadonnées :
    - indexation des métadonnées de préservation et mise à jour des journaux du cycle de vie concernés ;
    - sauvegarde des métadonnées et des journaux de cycle de vie sur les offres de stockage ;
- dans le cas d’une génération de binaire, mise à jour et **alimentation du registre des fonds** ;
- création d’un **rapport**, faisant état de l’opération de préservation ayant été exécutée.

À l’issue du traitement, l’opération peut aboutir aux statuts suivants :

Statut	Motifs <sup>37</sup>
Succès	Opération réalisée sans rencontrer de problèmes particuliers.
Avertissement	- Opération réalisée, sans qu’aucun traitement ne soit effectué pour cause de : <ul style="list-style-type: none"> <li>• lot d’archives sélectionné n’ayant pas d’objets numériques à traiter ;</li> <li>• scénario de préservation ne traitant pas les formats présents dans les objets à traiter.</li> </ul> - Opération réalisée, avec un traitement partiel effectué, dans le cas où une partie des objets traités sont en erreur.
Échec	- Taille maximale dépassée pour au moins un objet ; - Temps maximal de traitement dépassé pour au moins un objet ; - Action présente dans le scénario de préservation non gérée par le griffon utilisé.
Erreur technique <sup>38</sup>	Indisponibilité du griffon dans la solution logicielle Vitam.

préservation ».

37 La liste de motifs en cas d’avertissement, d’échec ou d’erreur technique n’est pas exhaustive.

38 Dans ce cas précis, l’opération peut également demeurer « En cours » ou s’arrêter brusquement, sans qu’aucun message ne soit

**Point d’attention :** le processus de préservation est **identique et unique**, quelle que soit l’action de préservation effectuée (analyse, génération de binaire, réidentification de format, extraction de métadonnées). De fait, certaines tâches sont automatiquement effectuées, mais non exécutées, en fonction des actions de préservation déclarées par le scénario de préservation. Par exemple, pour une opération d’analyse de formats, le processus affiche systématiquement une étape d’alimentation du registre des fonds, même si cette dernière n’est pas exécutée.

#### 4.1.2.3. Résultats d’une opération de préservation

Quelle que soit l’action de préservation effectuée, la solution logicielle Vitam produit un rapport listant les objets ayant fait l’objet de l’opération, ainsi que, pour chacun d’eux, le résultat de cette opération<sup>39</sup>.

En plus de ce rapport, la solution logicielle Vitam met en œuvre des mécanismes variant en fonction de l’action de préservation qui a été lancée :

	Création d’un objet	Mise à jour du groupe d’objets techniques	Mise à jour de l’unité archivistique	Mise à jour du registre des fonds	Rapport
<b>Génération de binaire</b>	oui	oui	non	oui	oui
<b>Identification simple (sans réidentification)</b>	non	non	non	non	oui
<b>Réidentification</b>	non	oui	non	non	oui
<b>Analyse</b>	non	non	non	non	oui
<b>Extraction de métadonnées</b>	non	oui	oui	non	oui

En résumé :

- à l’issue d’une génération de binaire, la solution logicielle Vitam met à jour les groupes d’objets techniques concernés en leur ajoutant un nouvel objet. Cette mise à jour est journalisée dans le journal du cycle de vie des groupes d’objets techniques concernés. Le registre des fonds est également mis à jour :
  - pour un service producteur donné, le nombre total d’objets est mis à jour, tandis que le nombre total de groupes d’objets techniques ne change pas,
  - une nouvelle opération est ajoutée, comptabilisant, pour un service producteur donné, le nombre d’objets créés dans le cadre de cette opération ;
- à l’issue d’une extraction de métadonnées, la solution logicielle Vitam est amenée à mettre à

précisé dans le journal des opérations.

<sup>39</sup> Pour plus d’informations sur l’opération de préservation, consulter le document *Modèle de workflow*, chapitre 11.I « Workflow de préservation ».

jour :

- soit les métadonnées des unités archivistiques associées aux objets dont les métadonnées sont extraites (*service non implémenté*),
- soit les métadonnées des groupes d'objets techniques,
- soit les deux (*service non implémenté*).

L'opération s'accompagne d'une journalisation dans les journaux de cycle de vie des unités archivistiques et/ou des groupes d'objets techniques concernés, en fonction du type de métadonnées qui est alimenté par cette opération d'extraction de métadonnées.

Une action d'analyse est uniquement accompagnée d'un rapport, associé à l'opération de préservation générée dans le journal des opérations.

### 4.1.3. Conseils de mise en œuvre

À l'issue de cette phase de réalisation de fonctionnalités concernant la préservation, l'équipe projet Vitam est en mesure de fournir quelques recommandations de mise en œuvre.

#### 4.1.3.1. Quel accès à la gestion de la préservation ?

##### *4.1.3.1.1. Gestion des droits*

En fonction de l'organisation adoptée, les opérations de préservation seront effectuées :

- soit par l'ensemble des utilisateurs de la solution logicielle Vitam,
- soit par un nombre restreint d'utilisateurs, dotés de droits élargis.

Un administrateur fonctionnel et/ou technique a accès à l'ensemble des services relatifs à la préservation.

**Point d'attention** : la génération de nouveaux binaires pouvant amener à créer de nouveaux objets sur tout ou partie des fonds conservés dans la solution, il est recommandé de proposer son utilisation à des utilisateurs avertis, en mesure de cibler des périmètres cohérents, méritant de faire l'objet d'une génération de binaires, et de planifier régulièrement des opérations de conversion.

En termes d'accès applicatifs :

- un système d'information archivistique a vocation à pouvoir proposer des services de préservation et à avoir accès, de fait, à la préservation ;
- une application versante et/ou accédante n'a pas forcément vocation à avoir accès à ces services.

##### *4.1.3.1.2. Restitution sur une IHM*

La solution logicielle Vitam propose de lancer les opérations de préservation depuis l'IHM démo en procédant comme suit :

- sélection d'unités archivistiques suivant des critères associés à celles-ci ;
- mise au panier de ces unités archivistiques ;

- lancement d’une opération de préservation depuis le panier.

S’il y a conception d’écrans, il est conseillé de privilégier des interfaces propres à la préservation, permettant :

- la sélection d’objets devant faire l’objet d’une opération de préservation à partir de critères aussi bien relatifs aux objets qu’aux unités archivistiques :
  - type MIME,
  - PUID,
  - usages,
  - service producteur,
  - type,
  - profil d’unité archivistique,
  - etc.
- partant de cette sélection, le paramétrage de l’opération de préservation, en fonction des critères suivant :
  - usage original des objets,
  - version des objets concernés,
  - le cas échéant, usage cible,
  - scénario de préservation.
- le lancement de l’opération.

#### 4.1.3.2. Comment utiliser et gérer la préservation ?

Intitulé	Description	Niveau de recommandation
<b>Gestion des droits associés à la préservation</b>		
Accès aux fonctionnalités de préservation par des administrateurs fonctionnel et/ou technique	Un administrateur fonctionnel et/ou technique a accès à l’ensemble des services relatifs à la préservation.	Obligatoire
Accès aux fonctionnalités de préservation par des archivistes	<p>En fonction de l’organisation adoptée, les opérations de préservation seront lancées :</p> <ul style="list-style-type: none"> <li>• soit par l’ensemble des utilisateurs de la solution logicielle Vitam,</li> <li>• soit par un nombre restreint d’utilisateurs, dotés de droits élargis.</li> </ul> <p>La génération de nouveaux binaires pouvant amener à créer de nouveaux objets sur tout ou partie des fonds conservés dans la solution, il est recommandé de proposer son utilisation à des <b>utilisateurs avertis et formés</b>, en mesure de cibler des périmètres cohérents, méritant de faire l’objet d’une génération de binaires, et de planifier régulièrement des opérations de</p>	Recommandé



	conversion	
Accès aux fonctionnalités de préservation par un SIA	Un système d'information archivistique a vocation à pouvoir proposer des services de préservation et à avoir accès, de fait, à la préservation.	Recommandé
Accès aux fonctionnalités de préservation par une application versante et/ou accédante	Une application versante et/ou accédante n'a pas vocation à avoir accès à ces services, à moins de nécessiter la mise en place de services liés à la préservation. Il est alors recommandé de pré-paramétrer les fonctionnalités de préservation de manière à ce qu'elles soient lancées au moyen de batchs lancés automatiquement, lors de la mise en œuvre de l'application.	Recommandé
<b>Restitution sur une IHM</b>		
Gestion de la préservation	S'il y a conception d'écrans, il est conseillé de privilégier des interfaces propres à la préservation, permettant : <ul style="list-style-type: none"> <li>• la sélection d'objets devant faire l'objet d'une opération de préservation à partir de critères aussi bien relatifs aux objets qu'aux unités archivistiques : type MIME, PUID, usages, service producteur, type, profil d'unité archivistique, etc.</li> <li>• partant de cette sélection, le paramétrage de l'opération de préservation, en fonction des critères suivant : <ul style="list-style-type: none"> <li>◦ usage original des objets,</li> <li>◦ version des objets concernés,</li> <li>◦ le cas échéant, usage cible,</li> <li>◦ scénario de préservation.</li> </ul> </li> <li>• le lancement de l'opération.</li> </ul>	Conseillé
<b>Analyse de formats</b>		
Vérification de la validité de formats	La solution logicielle Vitam intègre actuellement quatre griffons permettant d'analyser des formats : <ul style="list-style-type: none"> <li>• le griffon JHOVE qui permet de valider un nombre restreint de formats. Avec ce griffon, il n'est possible que de vérifier la validité des formats suivants : AIFF, ASCII, GIF, HTML, JPEG, JPEG 2000, PDF, TIFF, UTF-8, WAVE, XML, GZIP, PNG, WARC ;</li> <li>• le griffon ImageMagick qui permet de valider environ 200 formats de fichier<sup>40</sup> ;</li> <li>• le griffon ODFValidation qui permet de valider les formats de type Open Document ;</li> <li>• le griffon VeraPDF qui permet d'analyser les objets</li> </ul>	Optionnel

40 Une liste des formats gérés par ImageMagick est disponible à la page suivante : <https://imagemagick.org/script/formats.php> (Lien consulté le : 14 février 2019).

	<p>numériques au format PDF/A.</p> <p>En fonction des besoins, il est possible d’installer et d’utiliser l’un ou l’ensemble de ces griffons.</p>	
Vérification manuelle de la validité des formats sur un périmètre restreint d’objets	<p>La solution logicielle Vitam permet de vérifier la validité des formats sur tout ou partie des objets qu’elle conserve.</p> <p>Il est recommandé de procéder à une opération de ce type sur un périmètre restreint, le cas échéant faisant partie intégrante d’un programme complet de vérification de la validité des formats, afin de :</p> <ul style="list-style-type: none"> <li>• ne pas impacter les performances de la solution logicielle Vitam ;</li> <li>• avoir un temps de traitement court et la production d’un rapport d’opération rapidement disponible.</li> </ul>	Recommandé
Vérification manuelle de la validité des formats sur l’ensemble des objets conservés	<p>La solution logicielle Vitam permet de vérifier la validité des formats sur tout ou partie des objets qu’elle conserve.</p> <p>Néanmoins, il n’est pas recommandé de lancer une opération de ce type sur l’ensemble des objets conservés, car cette action pourrait avoir un impact sur :</p> <ul style="list-style-type: none"> <li>• les performances de la solution logicielle Vitam ;</li> <li>• le temps de traitement de l’opération. Cette dernière pourrait, en effet, durer plusieurs jours.</li> </ul>	Non recommandé
<b>Génération d’objets binaires / Conversion de formats</b>		
Conversion manuelle de formats	<p>La solution logicielle Vitam intègre actuellement trois griffons permettant de générer des objets :</p> <ul style="list-style-type: none"> <li>• le griffon ImageMagick qui permet de convertir 200 types de formats de fichier<sup>41</sup> ;</li> <li>• le griffon LibreOffice ;</li> <li>• le griffon Tesseract.</li> </ul> <p>Dès lors, pour chaque outil, il n’est possible que de convertir les formats que gère cet outil.</p>	Recommandé
Conversion de formats sur un périmètre restreint d’objets	<p>La solution logicielle Vitam permet de convertir des formats sur tout ou partie des objets qu’elle conserve.</p> <p>Il est recommandé de procéder à une opération de ce type sur un périmètre restreint, le cas échéant faisant partie intégrante d’un programme complet de conversion de formats, afin de :</p> <ul style="list-style-type: none"> <li>• ne pas impacter les performances de la solution logicielle Vitam ;</li> <li>• avoir un temps de traitement court et la production d’un rapport d’opération rapidement disponible.</li> </ul>	Recommandé
Conversion de formats sur	<p>La solution logicielle Vitam permet de convertir des formats sur tout ou partie des objets qu’elle conserve.</p>	Non recommandé

41 Une liste des formats gérés par ImageMagick est disponible à la page suivante : <https://imagemagick.org/script/formats.php> (Lien consulté le : 14 février 2019).

<p>l'ensemble des objets conservés</p>	<p>Néanmoins, il n'est pas recommandé de lancer une opération de ce type sur l'ensemble des objets conservés, car cette action pourrait avoir un impact sur :</p> <ul style="list-style-type: none"> <li>• les performances de la solution logicielle Vitam ;</li> <li>• le stockage disponible ;</li> <li>• le temps de traitement de l'opération. Cette dernière pourrait, en effet, durer plusieurs jours.</li> </ul>	
<p>Conversion systématique des formats</p>	<p>Dans le cadre de campagne de conversion d'un à plusieurs formats sur l'ensemble des fonds conservés dans la solution logicielle Vitam ou sur une partie importante d'entre eux, il est recommandé de programmer un batch, permettant de mener cette opération en fond de tâche, durant des heures de faible utilisation du système (en soirée ou durant le week-end).</p>	<p>Recommandé</p>
<p>Conversion d'un original numérique vers une nouvelle version d'original numérique</p>	<p>La conversion d'un original numérique vers une nouvelle version d'original numérique peut intervenir à plusieurs occasions :</p> <ul style="list-style-type: none"> <li>• Le format de l'original numérique est considéré comme obsolète et on lance une campagne de migration de format ;</li> <li>• Le format de l'original numérique n'est pas valide et, on souhaite générer un nouvel objet valide par rapport à un nouveau format ;</li> <li>• Le format de l'original numérique est inconnu, non identifié, et on souhaite générer un nouvel objet dont le format est identifié.</li> </ul>	<p>Recommandé</p>
<p>Conversion d'un original numérique vers une version de diffusion</p>	<p>La génération d'une version de diffusion depuis un original numérique peut intervenir pour des groupes d'objets ne contenant que des originaux numériques et pas de version de diffusion et à des fins de mise à disposition d'une version de diffusion, identifiée comme tel.</p> <p>Cette version de diffusion peut, notamment dans le cas des archives figurées et audiovisuelles, avoir des caractéristiques différentes par rapport à la version originale :</p> <ul style="list-style-type: none"> <li>• format différent,</li> <li>• volumétrie inférieure,</li> <li>• définition plus basse, de qualité moindre,</li> <li>• le cas échéant, présence d'un filigrane.</li> </ul>	<p>Recommandé</p>
<p>Conversion d'un original numérique vers une vignette</p>	<p>La génération d'une vignette depuis un original numérique peut intervenir pour des groupes d'objets ne contenant que des originaux numériques et pas de version de diffusion.</p> <p>Elle peut s'avérer nécessaire dans le cas de la mise en ligne de documents figurés sur un portail. Si ce dernier ne gère pas nativement la création de vignettes et en nécessite, il est recommandé de procéder à cette opération de préservation.</p> <p>Cette vignette aura des caractéristiques différentes par rapport à la version originale :</p>	<p>Recommandé</p>

	<ul style="list-style-type: none"> <li>• format éventuellement différent,</li> <li>• volumétrie inférieure,</li> <li>• taille inférieure,</li> <li>• le cas échéant, définition plus basse, de qualité moindre.</li> </ul>	
Conversion d'une version de diffusion vers une vignette	<p>La génération d'une vignette depuis une version de diffusion peut intervenir pour des groupes d'objets contenant seulement une version de diffusion ou contenant, en plus de l'original numérique, une version de diffusion.</p> <p>Elle peut s'avérer nécessaire dans le cas de la mise en ligne de documents figurés sur un portail. Si ce dernier ne gère pas nativement la création de vignettes et en nécessite, il est recommandé de procéder à cette opération de préservation.</p> <p>La vignette sera une copie dégradée d'une copie probablement déjà dégradée de l'original.</p>	Recommandé
Conversion d'une version de diffusion, une vignette ou un texte brut vers un original numérique	<p>La génération d'un original numérique depuis une version de diffusion, une vignette ou un texte brut n'est pas recommandée, dans la mesure où la version de diffusion, la vignette ou le texte brut sont l'émanation d'un original numérique, et non le contraire.</p>	Non recommandé
Conversion d'une version de diffusion, une vignette ou un texte brut vers un original numérique	<p>La génération d'un original numérique depuis une version de diffusion, une vignette ou un texte brut peut être envisagée et intervenir pour des groupes d'objets ne contenant pas d'original numérique et dont on considère qu'un des objets peut être considéré comme un original numérique, à défaut d'en avoir un dès l'origine.</p> <p>Par exemple, on peut considérer et convertir en originaux numériques des fichiers au format JPEG, transférés dans la solution logicielle Vitam comme objets de diffusion, si on sait que les véritables originaux numériques ne seront pas versés par le service producteur ou ont été détruits.</p>	Envisageable
<b>Identification / réidentification de formats</b>		
Identification manuelle de formats	<p>La solution logicielle Vitam intègre actuellement le griffon Siegfried qui permet d'identifier les formats référencés dans le référentiel des formats.</p>	Recommandé
Identification de formats sur un périmètre restreint d'objets	<p>La solution logicielle Vitam permet d'identifier des formats sur tout ou partie des objets qu'elle conserve.</p> <p>Il est recommandé de procéder à une opération de ce type sur un périmètre restreint, le cas échéant faisant partie intégrante d'un programme complet d'identification de formats, afin de :</p> <ul style="list-style-type: none"> <li>• ne pas impacter les performances de la solution logicielle Vitam ;</li> <li>• avoir un temps de traitement court et la production d'un rapport d'opération rapidement disponible.</li> </ul>	Recommandé

<p>Identification de formats sur l'ensemble des objets conservés</p>	<p>La solution logicielle Vitam permet d'identifier des formats sur tout ou partie des objets qu'elle conserve.</p> <p>Néanmoins, il n'est pas recommandé de lancer une opération de ce type sur l'ensemble des objets conservés, car cette action pourrait avoir un impact sur :</p> <ul style="list-style-type: none"> <li>• les performances de la solution logicielle Vitam ;</li> <li>• le stockage disponible ;</li> <li>• le temps de traitement de l'opération. Cette dernière pourrait, en effet, durer plusieurs jours.</li> </ul>	<p>Non recommandé</p>
<p>Réidentification de formats</p>	<p>Il est recommandé de lancer des opérations de réidentification de formats à l'occasion des mises à jour du référentiel des formats, afin de vérifier, notamment, si certains fichiers, entrés dans le système sans être identifiés, le sont avec le nouveau référentiel des formats.</p>	<p>Recommandé</p>
<p><b>Extraction de métadonnées</b></p>		
<p>Extraction manuelle de métadonnées</p>	<p>La solution logicielle Vitam intègre actuellement le griffon ImageMagick qui permet d'extraire des métadonnées internes issues d'environ 200 types de formats de fichier<sup>42</sup>.</p> <p>Dès lors, il n'est possible que d'extraire les métadonnées depuis des objets numériques dont le format est géré par cet outil.</p>	<p>Recommandé</p>
<p>Extraction de métadonnées internes sur un périmètre restreint d'objets</p>	<p>La solution logicielle Vitam permet d'extraire des métadonnées internes sur tout ou partie des objets qu'elle conserve.</p> <p>Il est recommandé de procéder à une opération de ce type sur un périmètre restreint, le cas échéant faisant partie intégrante d'un programme complet d'extraction de métadonnées, afin de :</p> <ul style="list-style-type: none"> <li>• ne pas impacter les performances de la solution logicielle Vitam ;</li> <li>• avoir un temps de traitement court et la production d'un rapport d'opération rapidement disponible.</li> </ul>	<p>Recommandé</p>
<p>Extraction de métadonnées internes sur l'ensemble des objets conservés</p>	<p>La solution logicielle Vitam permet d'extraire des métadonnées internes sur tout ou partie des objets qu'elle conserve.</p> <p>Néanmoins, il n'est pas recommandé de lancer une opération de ce type sur l'ensemble des objets conservés, car cette action pourrait avoir un impact sur :</p> <ul style="list-style-type: none"> <li>• les performances de la solution logicielle Vitam ;</li> <li>• le stockage disponible ;</li> <li>• le temps de traitement de l'opération. Cette dernière pourrait, en effet, durer plusieurs jours.</li> </ul>	<p>Non recommandé</p>
<p>Extraction de métadonnées internes systématique des</p>	<p>Dans le cadre de campagne d'extraction de métadonnées d'un à plusieurs formats sur l'ensemble des fonds conservés dans la solution logicielle Vitam ou sur une partie importante d'entre</p>	<p>Recommandé</p>

<sup>42</sup> Une liste des formats gérés par ImageMagick est disponible à la page suivante : <https://imagemagick.org/script/formats.php> (Lien consulté le : 14 février 2019).

formats	eux, il est recommandé de programmer un batch, permettant de mener cette opération en fond de tâche, durant des heures de faible utilisation du système (en soirée ou durant le week-end).	
Extraction de métadonnées internes	<p>L'extraction de métadonnées peut intervenir à plusieurs occasions :</p> <ul style="list-style-type: none"> <li>• avant la génération de nouvelle version d'objets, à des fins de préservation, dans la mesure où les métadonnées internes ne sont pas reportées dans le nouvel objet généré ;</li> <li>• pour rajouter automatiquement des informations aux éléments descriptifs associés à l'unité archivistique et/ou à l'objet technique, afin de les consulter et de les rechercher.</li> </ul>	Recommandé

## 4.2. Audit

La préservation numérique consiste en un ensemble de processus visant à maintenir un accès aux objets numériques sur le très long terme. En plus d'assurer un moyen de représenter ces objets dans le temps, cela implique de vérifier la disponibilité des objets dans la solution logicielle Vitam, ainsi que de maintenir leur intégrité.

Pour ce faire, la solution logicielle Vitam propose des fonctionnalités permettant d'assurer les audits suivants :

- audit portant sur les objets et leurs empreintes : audits d'existence et d'intégrité des objets,
- audit portant en plus sur la base de données et les journaux : audit de cohérence et audit correctif.

### 4.2.1. Audit d'existence et d'intégrité des objets

#### 4.2.1.1. Définition

**Dans l'audit d'existence des objets** : pour chaque objet défini dans une liste, la solution logicielle Vitam vérifie que :

- la liste des offres de stockage définie dans le groupe d'objets est bien la même que celle définie dans la stratégie de stockage ;
- tous les fichiers correspondant aux objets existent dans les offres spécifiées au moyen de la stratégie de stockage ;

**Dans l'audit d'intégrité des objets** : pour chaque objet défini dans une liste, la solution logicielle Vitam vérifie en plus que les empreintes des objets stockées et enregistrées dans la base de données sont identiques aux empreintes fournies par les offres de stockage.

Ces audits consistent en des actions d'évaluation, de vérification de l'existence et, le cas échéant, de l'intégrité d'une liste d'objets. Suite à cette analyse, une action corrective peut être lancée afin de rétablir l'accessibilité des objets absents ou corrompus, et ce à des fins de préservation numérique.

#### 4.2.1.2. Mécanismes mis en œuvre dans la solution logicielle Vitam

##### *4.2.1.1.1. Sélection de l'objet à auditer*

La solution logicielle Vitam permet de réaliser une opération d'audit :

- soit sur un tenant donné,
- soit sur un service producteur en particulier ;
- soit à partir d'une requête destinée à sélectionner un lot d'archives.

Dans les deux premiers cas, le nombre d'objets numériques à auditer n'est pas limité.

Dans le dernier cas, la requête est limitée par défaut à 10 000 objets, mais il est possible de modifier ce seuil pour obtenir un périmètre d'audit plus large.

Il est possible de réaliser deux audits :

- soit un audit vérifiant uniquement que l'objet existe bien,
- soit un audit vérifiant en même temps et l'existence de l'objet et son intégrité.

**Point d'attention :** en fonction du nombre d'objets conservés dans la solution logicielle Vitam, une opération d'audit peut prendre du temps et avoir un impact en termes de performance. Avant de lancer un audit, il est recommandé de bien prendre en considération le périmètre qui doit faire l'objet de cette opération.

##### *4.2.1.1.2. Opération d'audit*

Dans le cadre du processus d'audit d'un ensemble d'archives, suite à la réception d'une requête, la solution logicielle Vitam effectue les tâches et traitements de contrôles internes suivants<sup>43</sup> :

- **préparation** du traitement avec vérification des seuils de limitation des traitements portant sur les unités archivistiques et création de liste des groupes d'objets faisant l'objet de l'opération ;
- **action** d'audit, elle-même subdivisée en tâches :
  - vérification de l'existence des objets ;
  - le cas échéant, vérification de l'intégrité des objets ;
- **finalisation** de l'audit et création d'un **rapport**, faisant état de l'opération d'audit ayant été exécutée.

**Point d'attention :** le processus d'audit de l'intégrité des objets intègre une tâche de vérification de l'existence des objets.

À l'issue de l'audit, l'opération peut aboutir aux statuts suivants :

<b>Statut</b>	<b>Motifs</b>
Succès	Opération réalisée sans rencontrer de problèmes particuliers.
Avertissement	Opération réalisée, sans qu'aucun traitement ne soit effectué pour cause de lot d'archives

<sup>43</sup> Pour plus d'informations sur l'opération d'audit, consulter le document *Modèle de workflow*, chapitre 2.II « Workflow de l'audit de l'existence et de l'intégrité des fichiers ».

	sélectionné n'ayant pas d'objets numériques à traiter.
Échec	- au moins un objet est absent dans l'une des offres déclarées et par rapport au nombre de copies déclarées dans la stratégie de stockage ; - au moins une empreinte d'objet stocké et enregistré dans la base de données n'est pas identique à l'empreinte fournie par les espaces de stockage.
Erreur technique	- Indisponibilité du service dans la solution logicielle Vitam.

#### 4.2.1.1.2. Résultat d'une opération d'audit

Quelle que soit l'action d'audit effectuée, la solution logicielle Vitam produit systématiquement un rapport listant<sup>44</sup> :

- la liste des opérations d'entrée ayant fait l'objet de l'opération ;
- les objets ayant fait l'objet de l'opération en échec ;
- la liste d'éléments singuliers rencontrés (cas en avertissement).

#### 4.2.1.3. Conseils de mise en œuvre

##### 4.2.1.3.1. À quoi servent l'audit d'existence et l'audit d'intégrité des objets ?

L'audit d'existence des objets permet de vérifier la **présence effective** des objets préalablement transférés dans la solution logicielle Vitam. Il assure que ces objets sont toujours **disponibles** et qu'aucune de leurs copies, stockées dans différentes offres de stockage, conformément à la stratégie de stockage, n'a disparu des différentes offres suite à une erreur technique.

L'audit d'intégrité des objets permet de vérifier, outre l'existence des objets, que ces objets conservés dans la solution logicielle Vitam sont intègres, c'est-à-dire qu'**ils n'ont pas été modifiés** au sein de la solution logicielle Vitam.

##### 4.2.1.3.2. Quand et comment lancer un audit d'existence et un audit d'intégrité des objets ?

La solution logicielle Vitam permet de lancer des opérations d'audit d'existence et/ou d'intégrité :

- depuis l'IHM démo sur l'ensemble d'un tenant ou en sélectionnant un service producteur en particulier ;
- par l'API, au moyen d'une requête DSL.

Les raisons pouvant amener à réaliser un audit d'existence des objets sont les suivantes :

- audits ponctuels, par échantillon d'archives, afin de vérifier que les objets sont toujours conservés dans la solution logicielle Vitam ;
- audits sur des périmètres définis d'archives, faisant partie intégrante d'un programme complet, pluri-annuel, de vérification de l'existence d'objets ;
- audits initiés à l'occasion d'une élection, à des fins de récolement ;
- audits occasionnés par une évolution de stratégie de stockage ou d'offre de stockage.

<sup>44</sup> Pour plus d'informations sur l'opération d'audit, consulter le document *Modèle de workflow*, chapitre 2.III « Rapport d'audit ».



En plus de ces dernières, les raisons pouvant amener à réaliser un audit d'intégrité des objets sont les suivantes :

- prouver l'intégrité et l'authenticité d'un fonds en particulier, par la génération d'un rapport.

**Point d'attention :** Il est recommandé de procéder à une opération de ce type sur un périmètre restreint (service producteur par service producteur ou entrée par entrée), afin de :

- ne pas impacter les performances de la solution logicielle Vitam ;
- avoir un temps de traitement court et la production d'un rapport d'opération rapidement disponible.

#### 4.2.1.3.3. Comment utiliser l'audit d'existence et d'intégrité des objets ?

À l'issue de cette phase de réalisation de fonctionnalités concernant le référentiel Vitam, l'équipe projet Vitam est en mesure de fournir quelques recommandations de mise en œuvre :

Intitulé	Description	Niveau de recommandation
<b>Audit d'existence des objets</b>		
Vérification manuelle de l'existence d'objets sur un périmètre restreint d'objets	<p>La solution logicielle Vitam permet de vérifier l'existence d'objets sur tout ou partie des objets qu'elle conserve.</p> <p>Il est recommandé de procéder à une opération de ce type sur un périmètre restreint (service producteur par service producteur ou entrée par entrée), le cas échéant faisant partie intégrante d'un programme complet de vérification de l'existence d'objets, afin de :</p> <ul style="list-style-type: none"> <li>• ne pas impacter les performances de la solution logicielle Vitam ;</li> <li>• avoir un temps de traitement court et la production d'un rapport d'opération rapidement disponible.</li> </ul>	Recommandé
Vérification manuelle de l'existence d'objets sur l'ensemble des objets conservés	<p>La solution logicielle Vitam permet de vérifier l'existence d'objets sur tout ou partie des objets qu'elle conserve.</p> <p>Néanmoins, il n'est pas recommandé de lancer une opération de ce type sur l'ensemble des objets conservés, car cette action pourrait avoir un impact sur :</p> <ul style="list-style-type: none"> <li>• les performances de la solution logicielle Vitam ;</li> <li>• le temps de traitement de l'opération. Cette dernière pourrait, en effet, durer plusieurs jours.</li> </ul>	Non recommandé
Vérification systématique de l'existence d'objets	<p>Dans le cadre de campagne de vérification de l'existence d'objets sur l'ensemble des fonds conservés dans la solution logicielle Vitam ou sur une partie importante d'entre eux, il est recommandé de programmer un batch, permettant de mener cette opération, segmentée en petits lots d'archives, en fond de tâche, durant des heures de faible utilisation du</p>	Recommandé

	système (en soirée ou durant le week-end).	
Vérification systématique de l'existence d'objets entrants	<p>Dans le cas d'un audit systématique des objets transférés dans la solution logicielle Vitam, il est recommandé de programmer un batch auditant les entrées et permettant de mener cette opération :</p> <ul style="list-style-type: none"> <li>• quotidiennement (en soirée), afin de vérifier que les objets transférés dans la journée ont bien été stockés dans la solution logicielle Vitam ;</li> <li>• une fois par semaine (durant le week-end), afin de vérifier que les objets transférés dans la semaine ont bien été stockés dans la solution logicielle Vitam.</li> </ul>	Recommandé
<b>Audit d'intégrité des objets</b>		
Vérification manuelle de l'intégrité d'objets sur un périmètre restreint d'objets	<p>La solution logicielle Vitam permet de vérifier l'intégrité d'objets sur tout ou partie des objets qu'elle conserve. Il est recommandé de procéder à une opération de ce type sur un périmètre restreint (service producteur par service producteur ou entrée par entrée), le cas échéant faisant partie intégrante d'un programme complet de vérification de l'existence d'objets, afin de :</p> <ul style="list-style-type: none"> <li>• ne pas impacter les performances de la solution logicielle Vitam ;</li> <li>• avoir un temps de traitement court et la production d'un rapport d'opération rapidement disponible.</li> </ul>	Recommandé
Vérification manuelle de l'intégrité d'objets sur l'ensemble des objets conservés	<p>La solution logicielle Vitam permet de vérifier l'intégrité d'objets sur tout ou partie des objets qu'elle conserve. Néanmoins, il n'est pas recommandé de lancer une opération de ce type sur l'ensemble des objets conservés, car cette action pourrait avoir un impact sur :</p> <ul style="list-style-type: none"> <li>• les performances de la solution logicielle Vitam ;</li> <li>• le temps de traitement de l'opération. Cette dernière pourrait, en effet, durer plusieurs jours.</li> </ul>	Non recommandé
Vérification systématique de l'intégrité d'objets	<p>Dans le cadre de campagne de vérification de l'intégrité d'objets sur l'ensemble des fonds conservés dans la solution logicielle Vitam ou sur une partie importante d'entre eux, il est recommandé de programmer un batch, permettant de mener cette opération, segmentée en petits lots d'archives, en fond de tâche, durant des heures de faible utilisation du système (en soirée ou durant le week-end).</p>	Recommandé
Vérification systématique de l'intégrité d'objets entrants	<p>Dans le cas d'un audit systématique des objets transférés dans la solution logicielle Vitam, il est recommandé de programmer un batch auditant les entrées et permettant de mener cette opération :</p> <ul style="list-style-type: none"> <li>• quotidiennement (en soirée), afin de vérifier que</li> </ul>	Recommandé

	<p>les objets transférés dans la journée ont bien été stockés et sont intègres dans la solution logicielle Vitam ;</p> <ul style="list-style-type: none"> <li>• une fois par semaine (durant le week-end), afin de vérifier que les objets transférés dans la semaine ont bien été stockés et sont intègres dans la solution logicielle Vitam.</li> </ul>	
--	---	--

## 4.2.2. Audit de cohérence

### 4.2.2.1. Définition

Dans un **audit de cohérence**, la solution logicielle Vitam vérifie, pour chaque unité archivistique définie dans une liste, que :

- l’empreinte des métadonnées décrivant une unité archivistique est identique à celle qui est enregistrée dans le dernier journal sécurisé ;
- l’empreinte du journal du cycle de vie d’une unité archivistique est identique à celle qui est enregistrée dans le dernier journal sécurisé ;
- l’empreinte du groupe formé par les métadonnées et le journal du cycle de vie associés à une unité archivistique, stocké sur disque pour conservation et reconstruction, est identique à celle qui est enregistrée dans le dernier journal sécurisé ;
- si l’unité archivistique dispose d’un groupe d’objet technique :
  - pour chaque objet binaire :
    - l’empreinte de l’objet binaire de chaque offre est identique à celle du dernier journal sécurisé ;
    - l’empreinte de l’objet binaire enregistré en base est identique à celle du dernier journal sécurisé ;
  - l’empreinte des métadonnées décrivant un groupe d’objets techniques est identique à celle du dernier journal sécurisé ;
- l’empreinte du groupe formé par les métadonnées et le journal du cycle de vie associés à un groupe d’objets techniques, stocké sur disque pour conservation et reconstruction, est identique à celle qui est enregistrée dans le dernier journal sécurisé ;
  - l’empreinte du journal du cycle de vie d’un groupe d’objets techniques est identique à celle du dernier journal sécurisé.

Cet audit consiste en des actions d’évaluation, de vérification de l’existence et, le cas échéant, de l’intégrité d’une liste d’objets, de groupes d’objets techniques et d’unités archivistiques, ainsi que **de leur cohérence par rapport aux journaux sécurisés**. Suite à cette analyse, une action corrective peut être lancée afin de rétablir l’accessibilité des objets absents ou corrompus, et ce à des fins de préservation numérique.

#### 4.2.2.2. Mécanismes mis en œuvre dans la solution logicielle Vitam

##### 4.2.2.2.1. Sélection de l'objet à auditer

La solution logicielle Vitam permet de réaliser une opération d'audit de cohérence à partir d'un lot d'archives.

**Point d'attention :** en fonction du nombre d'objets conservés dans la solution logicielle Vitam, une opération d'audit peut prendre du temps et avoir un impact en termes de performance. Avant de lancer un audit, il est recommandé de bien prendre en considération le périmètre qui doit faire l'objet de cette opération.

##### 4.2.2.2.2. Opération d'audit

Dans le cadre du processus d'audit de cohérence d'un ensemble d'archives, suite à la réception d'une requête, la solution logicielle Vitam effectue les tâches et traitements de contrôles internes suivants<sup>45</sup> :

- **préparation** du traitement avec vérification des seuils de limitation des traitements portant sur les unités archivistiques et avec création de la liste faisant l'objet de l'opération ;
- **recupération des données** présentes dans la base de données ;
- extraction de la **liste des enregistrements** présents dans les derniers journaux sécurisés ;
- création d'un **rapport pour chaque entité vérifiée** : unité archivistique, objet ou groupe d'objets techniques ;
- **finalisation** de l'audit et création d'un **rapport**, faisant état de l'opération d'audit ayant été exécutée.

À l'issue de l'audit, l'opération peut aboutir aux statuts suivants :

Statut	Motifs
Succès	Opération réalisée sans rencontrer de problèmes particuliers.
Avertissement	Opération réalisée, sans qu'aucun traitement ne soit effectué pour cause de lot d'archives sélectionné n'ayant pas d'objets numériques à traiter.
Échec	<ul style="list-style-type: none"> <li>- l'empreinte des métadonnées décrivant une unité archivistique est différente de celle qui est enregistrée dans le dernier journal sécurisé ;</li> <li>- l'empreinte du journal du cycle de vie d'une unité archivistique est différente de celle qui est enregistrée dans le dernier journal sécurisé ;</li> <li>- l'empreinte du groupe formé par les métadonnées et le journal du cycle de vie associés à une unité archivistique, stocké sur disque pour conservation et reconstruction, est différente de celle qui est enregistrée dans le dernier journal sécurisé ;</li> <li>- si l'unité archivistique dispose d'un groupe d'objet technique :                             <ul style="list-style-type: none"> <li>• pour chaque objet binaire :                                     <ul style="list-style-type: none"> <li>○ l'empreinte de l'objet binaire de chaque offre est différente de celle du dernier journal sécurisé ;</li> <li>○ l'empreinte de l'objet binaire enregistré en base est différente de celle du</li> </ul> </li> </ul> </li> </ul>

<sup>45</sup> Pour plus d'informations sur l'opération d'audit, consulter le document *Modèle de workflow*, chapitre 2.IV « Workflow d'audit de cohérence des fichiers ».

	<p>dernier journal sécurisé ;</p> <ul style="list-style-type: none"> <li>• l’empreinte des métadonnées décrivant un groupe d’objets techniques est différente de celle du dernier journal sécurisé ;</li> </ul> <p>- l’empreinte du groupe formé par les métadonnées et le journal du cycle de vie associés à un groupe d’objets techniques, stocké sur disque pour conservation et reconstruction, est différente de celle qui est enregistrée dans le dernier journal sécurisé ;</p> <p>- l’empreinte du journal du cycle de vie d’un groupe d’objets techniques est différente de celle du dernier journal sécurisé.</p>
Erreur technique	- Indisponibilité du service dans la solution logicielle Vitam.

#### 4.2.2.2.3. Résultat d’une opération d’audit

Lors d’un audit de cohérence, la solution logicielle Vitam ne produit un rapport que dans les cas d’erreurs avérées<sup>46</sup>.

Chaque section du rapport correspond au résultat de l’audit de cohérence pour chaque objet ou groupe d’objets ou unité archivistique audité en erreur. On y trouve les informations suivantes :

- « identifier » : Identifiant de l’objet ou groupe d’objets ou unité archivistique audité.
- « status » : « KO » : le statut de l’opération (dans le cadre de larelease en cours, le statut sera systématiquement KO).
- « message » : message qui signale une incohérence entre les enregistrements des fichiers sécurisés, des offres de stockage et de la base de données.
- « objectType » : type de l’objet audité : objet ou groupe d’objets ou unité archivistique.
- « securedHash » : hash du journal sécurisé de l’unité archivistique, objet ou groupe d’objets.
- « offersHashes » : signatures de l’élément audité de type unit ou groupes d’objets techniques dans les offres de stockage.

En cas de succès ou d’avertissement le rapport reste vide.

#### 4.2.2.3. Conseils de mise en œuvre

##### 4.2.2.3.1. À quoi sert l’audit de cohérence ?

L’audit de cohérence permet de vérifier la **conformité** globale du système, puisqu’il croise empreintes des unités archivistiques et des groupes d’objets techniques, ainsi que de leurs journaux avec celles du journal sécurisé. Il offre une garantie supplémentaire en prouvant, au moyen de son rapport, que le système ou, du moins, le périmètre audité n’a pas subi d’altération.

##### 4.2.2.3.2. Quand et comment lancer un audit de cohérence ?

La solution logicielle Vitam permet de lancer des opérations d’audit de cohérence à partir d’un lot d’archives préalablement sélectionné.

Les raisons pouvant amener à réaliser un audit de cohérence sont les suivantes :

- prouver l’intégrité et l’authenticité d’un fonds en particulier, par la génération d’un rapport ;

<sup>46</sup> Pour plus d’informations sur l’opération d’audit, consulter le document *Modèle de workflow*, chapitre 2.V « Rapport d’audit de cohérence ».

- mesurer la conformité du système par rapport aux attendus de la norme AFNOR NF Z 42-013 ;
- audits ponctuels, associés à une procédure d’audit de certification ou de conformité ;
- fournir une évaluation objective de la qualité et de la fiabilité du système au moyen d’un rapport.

**Point d’attention :** Il est recommandé de procéder à une opération de ce type sur un périmètre restreint (service producteur par service producteur), afin de :

- ne pas impacter les performances de la solution logicielle Vitam ;
- avoir un temps de traitement court et la production d’un rapport d’opération rapidement disponible.

#### 4.2.2.3.3. Comment utiliser l’audit de cohérence ?

À l’issue de cette phase de réalisation de fonctionnalités concernant le référentiel Vitam, l’équipe projet Vitam est en mesure de fournir quelques recommandations de mise en œuvre :

Intitulé	Description	Niveau de recommandation
<b>Audit de cohérence</b>		
Vérification manuelle de la cohérence du système sur un périmètre restreint d’objets et des unités archivistiques	<p>La solution logicielle Vitam permet de vérifier la cohérence sur tout ou partie des objets et des unités archivistiques qu’elle conserve.</p> <p>Il est recommandé de procéder à une opération de ce type sur un périmètre restreint (service producteur par service producteur ou entrée par entrée), le cas échéant faisant partie intégrante d’un programme complet de vérification de l’existence d’objets et d’unités archivistiques, afin de :</p> <ul style="list-style-type: none"> <li>• ne pas impacter les performances de la solution logicielle Vitam ;</li> <li>• avoir un temps de traitement court et la production d’un rapport d’opération rapidement disponible.</li> </ul>	Recommandé
Vérification manuelle de la cohérence du système sur l’ensemble des objets et des unités archivistiques conservés	<p>La solution logicielle Vitam permet de vérifier la cohérence du système sur tout ou partie des objets et des unités archivistiques qu’elle conserve.</p> <p>Néanmoins, il n’est pas recommandé de lancer une opération de ce type sur l’ensemble des objets et des unités archivistiques conservés, car cette action pourrait avoir un impact sur :</p> <ul style="list-style-type: none"> <li>• les performances de la solution logicielle Vitam ;</li> <li>• le temps de traitement de l’opération. Cette dernière pourrait, en effet, durer plusieurs jours.</li> </ul>	Non recommandé
Vérification systématique de la	Dans le cadre de campagne de vérification de la cohérence du système sur l’ensemble des fonds conservés dans la solution	Recommandé

cohérence du système	logicielle Vitam ou sur une partie importante d’entre eux, il est recommandé de programmer un batch, permettant de mener cette opération en fond de tâche, durant des heures de faible utilisation du système (en soirée ou durant le week-end).	
Vérification systématique de la cohérence du système	<p>Dans le cas d’un audit systématique du système dans la solution logicielle Vitam, il est recommandé de programmer un batch auditant les dernières activités sur les objets et les unités archivistiques et permettant de mener cette opération :</p> <ul style="list-style-type: none"> <li>• quotidiennement (en soirée), afin de vérifier le maintien de la cohérence du système jour après jour ;</li> <li>• une fois par semaine (durant le week-end), afin de vérifier le maintien de la cohérence du système semaine après semaine.</li> </ul>	Recommandé

### 4.2.3. Audit correctif

#### 4.2.3.1. Définition

L’audit correctif comprend deux phases :

- un audit de cohérence :
  - pour chaque unité archivistique définie dans une liste, la solution logicielle Vitam vérifie que :
  - l’empreinte des métadonnées décrivant une unité archivistique est identique à celle qui est enregistrée dans le dernier journal sécurisé ;
  - l’empreinte du journal du cycle de vie d’une unité archivistique est identique à celle qui est enregistrée dans le dernier journal sécurisé ;
  - l’empreinte du groupe formé par les métadonnées et le journal du cycle de vie associés à une unité archivistique, stocké sur disque pour conservation et reconstruction, est identique à celle qui est enregistrée dans le dernier journal sécurisé ;
  - si l’unité archivistique dispose d’un groupe d’objet technique :
    - pour chaque objet binaire :
      - l’empreinte de l’objet binaire de chaque offre est identique à celle du dernier journal sécurisé ;
      - l’empreinte de l’objet binaire enregistré en base est identique à celle du dernier journal sécurisé ;
    - l’empreinte des métadonnées décrivant un groupe d’objets techniques est identique à celle du dernier journal sécurisé ;
  - l’empreinte du groupe formé par les métadonnées et le journal du cycle de vie associés à un groupe d’objets techniques, stocké sur disque pour conservation et reconstruction, est identique à celle qui est enregistrée dans le dernier journal sécurisé ;
    - l’empreinte du journal du cycle de vie d’un groupe d’objets techniques est identique à celle du dernier journal sécurisé ;
- sur la base de l’audit de cohérence, correction des anomalies rencontrées (par exemple,

rétablissement d'un objet dans son état antérieur).

Cet audit consiste en des actions d'évaluation, de vérification de l'existence et, le cas échéant, de l'intégrité d'une liste d'objets, de groupes d'objets techniques et d'unités archivistiques, ainsi que **de leur cohérence par rapport aux journaux sécurisés**. Suite à cette analyse, une action corrective est lancée afin de rétablir l'accessibilité des objets absents ou corrompus, et ce à des fins de préservation numérique.

#### 4.2.3.2. Mécanismes mis en œuvre dans la solution logicielle Vitam

##### *4.2.3.2.1. Sélection de l'objet à auditer*

La solution logicielle Vitam permet de réaliser une opération d'audit correctif à partir d'une opération d'audit de cohérence sur un tenant donné. De fait, il faut préalablement avoir réalisé un audit de cohérence avant de lancer un audit correctif.

**Point d'attention :** en fonction du nombre d'objets conservés dans la solution logicielle Vitam, une opération d'audit peut prendre du temps et avoir un impact en termes de performance. Avant de lancer un audit, il est recommandé de bien prendre en considération le périmètre qui doit faire l'objet de cette opération.

##### *4.2.2.3.2. Opération d'audit*

Dans le cadre du processus d'audit correctif d'un ensemble d'archives, suite à la réception d'une requête, la solution logicielle Vitam effectue les tâches et traitements de contrôles internes suivants<sup>47</sup> :

- **préparation** du traitement avec vérification des seuils de limitation des traitements portant sur les unités archivistiques et avec création de la liste faisant l'objet de l'opération ;
- **recupération des données** présentes dans la base de données ;
- extraction de la **liste des enregistrements** présents dans les derniers journaux sécurisés ;
- création d'un **rapport pour chaque entité vérifiée** : unité archivistique, objet ou groupe d'objets techniques ;
- **finalisation** de l'audit et création d'un **rapport**, faisant état de l'opération d'audit ayant été exécutée ;
- **correction de l'enregistrement** pour un objet, un groupe d'objets ou une unité archivistique donné(e) défaillant(e) ;
- **finalisation** des corrections des signatures sur les données du/des fichier(s) auprès de la base de données ou d'une offre de stockage valide.

À l'issue de l'audit, l'opération peut aboutir aux statuts suivants :

<b>Statut</b>	<b>Motifs</b>
Succès	Opération réalisée sans rencontrer de problèmes particuliers.

<sup>47</sup> Pour plus d'informations sur l'opération d'audit, consulter le document *Modèle de workflow*, chapitre 2.VI « Workflow de l'audit correctif ».



Avertissement	Opération réalisée, sans qu’aucun traitement ne soit effectué pour cause de lot d’archives sélectionné n’ayant pas d’objets numériques à traiter.
Échec	Il y a plus d’un objet, d’un groupe d’objets techniques ou d’une unité archivistique à corriger.
Erreur technique	Indisponibilité du service dans la solution logicielle Vitam.

#### 4.2.3.2.3. Résultat d’une opération d’audit

Lors d’un audit correctif, la solution logicielle Vitam ne produit un rapport que dans les cas d’erreurs avérées<sup>48</sup>.

Chaque section du rapport correspond au résultat de l’audit pour chaque objet ou groupe d’objets ou unité archivistique auditée en erreur. On y trouve les informations suivantes :

- « identifier » : Identifiant de l’objet ou groupe d’objets ou unité archivistique auditée.
- « status » : « KO » : le statut de l’opération (dans le cadre de larelease en cours, le statut sera systématiquement KO).
- « message » : message qui signale une incohérence entre les enregistrements des fichiers sécurisés, des offres de stockage et de la base de données.
- « objectType » : type de l’objet auditée : objet ou groupe d’objets ou unité archivistique.
- « securedHash » : hash du journal sécurisé de l’unité archivistique, objet ou groupe d’objets.
- « offersHashes » : signatures de l’élément auditée de type unit ou groupes d’objets techniques dans les offres de stockage.

En cas de succès ou d’avertissement le rapport reste vide.

#### 4.2.3.3. Conseils de mise en œuvre

##### 4.2.3.3.1. À quoi sert l’audit correctif ?

L’audit correctif permet de vérifier la **conformité** globale du système, en procédant à un nouvel audit de cohérence. Mais il va au-delà en **rétablissant les éléments manquants** trouvés par cet audit. De fait, il garantit que la solution logicielle Vitam conserve toujours les mêmes archives.

**Point d’attention :** La solution logicielle Vitam ne rétablit que l’objet présent sur une offre non corrompue. En revanche, elle ne rétablit pas les incohérences présentes dans la base de données.

##### 4.2.3.3.2. Quand et comment lancer un audit correctif ?

Un audit correctif ne peut être lancé qu’à la suite d’un audit de cohérence dont le rapport contient au moins une anomalie. En d’autres termes, l’audit correctif **a pour but de corriger les erreurs de l’audit de cohérence**, préalablement lancé.

Si l’audit de cohérence n’a montré aucune anomalie, il n’est pas nécessaire de lancer un audit

<sup>48</sup> Pour plus d’informations sur l’opération d’audit, consulter le document *Modèle de workflow*, chapitre 2.V « Rapport d’audit de cohérence ».

correctif.

L’audit correctif s’effectue depuis l’IHM recette mise à disposition.

**Point d’attention :** Il est recommandé de procéder à une opération de ce type sur un périmètre restreint (service producteur par service producteur), afin de :

- ne pas impacter les performances de la solution logicielle Vitam ;
- avoir un temps de traitement court et la production d’un rapport d’opération rapidement disponible.

#### 4.2.3.3. Comment utiliser l’audit correctif ?

À l’issue de cette phase de réalisation de fonctionnalités concernant le référentiel Vitam, l’équipe projet Vitam est en mesure de fournir quelques recommandations de mise en œuvre :

Intitulé	Description	Niveau de recommandation
<b>Audit correctif</b>		
Lancement d’un audit correctif	Un audit correctif ne peut être réalisé qu’à la suite d’un audit de cohérence.	Obligatoire
Vérification manuelle et corrections des incohérences du système sur un périmètre restreint d’objets et d’unités archivistique	La solution logicielle Vitam permet de corriger les incohérences du système sur tout ou partie des objets qu’elle conserve. Il est recommandé de procéder à une opération de ce type si un audit de cohérence a préalablement mis en évidence des anomalies.	Recommandé
Vérification manuelle de la cohérence du système sur l’ensemble des objets et des unités archivistiques conservés et correction des anomalies	La solution logicielle Vitam permet de vérifier la cohérence du système sur tout ou partie des objets et des unités archivistiques qu’elle conserve et de corriger les anomalies rencontrées. Néanmoins, il n’est pas recommandé de lancer une opération de ce type sur l’ensemble des objets et des unités archivistiques conservés, car : <ul style="list-style-type: none"> <li>• cette action pourrait avoir un impact sur : <ul style="list-style-type: none"> <li>◦ les performances de la solution logicielle Vitam ;</li> <li>◦ le temps de traitement de l’opération. Cette dernière pourrait, en effet, durer plusieurs jours ;</li> </ul> </li> <li>• s’il y a plusieurs anomalies, l’audit correctif ne fonctionnera pas.</li> </ul>	Non recommandé
Vérification et correction systématiques des objets et des unités	Dans le cadre de campagne de vérification de la cohérence du système sur l’ensemble des fonds conservés dans la solution logicielle Vitam ou sur une partie importante d’entre	Recommandé

archivistiques	eux, suivie d'une correction automatique des objets, il est recommandé de programmer un batch lançant un audit de cohérence et, suivant les résultats de ce dernier, un batch lançant à la suite un audit correctif traitant les anomalies une à une. Le batch permet de mener cette opération en fond de tâche, durant des heures de faible utilisation du système (en soirée ou durant le week-end).	
----------------	--	--

## Annexe 1 : exemples de référentiels de préservation

*Nota bene* : les cas présentés ci-dessous sont des exemples fictifs et visent simplement à vérifier la bonne mise en œuvre des mécanismes relatifs aux habilitations dans la solution logicielle Vitam.

### A. Griffons

```
[
  {
    "Identifiant": "GRI-000001",
    "Name": "Griffon 1 : ImageMagick",
    "Description": "ImageMagick est un outil de validation et de conversion de fichiers image. Il a avant tout été développé pour créer, modifier, composer et convertir des images. Il est capable de traiter 200 types de formats de fichiers – dont PNG, JPEG, GIF, HEIC, TIFF, DPX, EXR, WebP, Postscript, PDF, et SVG.",
    "ExecutableName": "imagemagick-griffin",
    "ExecutableVersion": "V1.0.0"
  },
  {
    "Identifiant": "GRI-000002",
    "Name": "Griffon 2 : Griffon Jhove",
    "Description": "JHOVE est un outil d'identification, de validation et de conversion de fichiers. Il traite les formats suivants : AIFF, ASCII, GIF, HTML, JPEG, JPEG 2000, PDF, TIFF, UTF-8, WAVE, XML, GZIP, PNG, WARC.",
    "ExecutableName": "jhove-griffin",
    "ExecutableVersion": "V1.0.0"
  }
]
```

### B. Scénarios de préservation

```
{
  "Identifiant": "PSC-000001",
  "Name": "Transformation en GIF (fmt/43)",
  "Description": "Ce scénario, appelant le griffon ImageMagick, permet de convertir des fichiers dont le PUID est fmt/43 en fichiers au format GIF, et prenant la forme de vignettes de 100px x 100px",
  "CreationDate": "2018-11-16T15:55:30.721",
  "ActionList": [
    "GENERATE"
  ],
  "GriffinByFormat": [
    {
      "FormatList": ["fmt/43"],
      "GriffinIdentifiant": "GRI-000001",
      "Timeout": 2000,
    }
  ]
}
```

```
"MaxSize": 10000000,  
"Debug":true,  
"ActionDetail": [  
  {  
    "Type": "GENERATE",  
    "Values": {  
      "Extension": "GIF",  
      "Args": [  
        "-thumbnail",  
        "100x100"  
      ]  
    }  
  }  
]
```

## Annexe 2 : Messages d’erreur

*Nota bene* : les cas présentés ci-dessous n’ont pas vocation à être exhaustif.

	Message retourné par la solution logicielle Vitam	Explication	Résolution
<b>Import ou mise à jour d’un référentiel des griffons</b>			
1	{ "ErrorDetail" : "Duplicate griffin : 'GRIFFIN1 '" }	L'utilisateur a importé un référentiel contenant deux griffons ayant le même identifiant.	Modifier l'identifiant de l'un des deux griffons dont l'identifiant est égal à « GRIFFIN1 ».
2	{ "ErrorDetail" : "Invalid griffin for : 'null' : ['identifiant' : this field can not be empty]" }	Deux possibilités : <ul style="list-style-type: none"> <li>soit un des griffons ne contient pas le champ obligatoire « Identifiant » ;</li> <li>soit un des griffons contient un champ « Identifiant » sans valeur, alors que c'est obligatoire.</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>soit ajouter un champ « Identifiant » au griffon concerné ;</li> <li>soit rajouter une valeur dans le champ « Identifiant » du griffon concerné.</li> </ul>
3	{ "ErrorDetail" : "Invalid griffin for : 'GRIFFIN1' : ['name' : this field can not be empty]" }	Deux possibilités pour le griffon dont l'identifiant est « GRIFFIN1 » : <ul style="list-style-type: none"> <li>soit il ne contient pas le champ obligatoire « Name » ;</li> <li>soit il contient un champ « Name » sans valeur, alors que c'est obligatoire.</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>soit ajouter un champ « Name » au griffon concerné ;</li> <li>soit rajouter une valeur dans le champ « Name » du griffon concerné.</li> </ul>
4	{ "ErrorDetail" : "Invalid griffin for : 'GRIFFIN1' : ['ExecutableName' : this field can not be empty]" }	Deux possibilités pour le griffon dont l'identifiant est « GRIFFIN1 » : <ul style="list-style-type: none"> <li>soit il ne contient pas le champ obligatoire « ExecutableName » ;</li> <li>soit il contient un champ « ExecutableName » sans valeur, alors que c'est obligatoire.</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>soit ajouter un champ « ExecutableName » au griffon concerné ;</li> <li>soit rajouter une valeur dans le champ « ExecutableName » du griffon concerné.</li> </ul>
5	{ "ErrorDetail" : "Invalid griffin for : 'GRIFFIN1' : ['ExecutableVersion' : this field can not be empty]" }	Deux possibilités pour le griffon dont l'identifiant est « GRIFFIN1 » : <ul style="list-style-type: none"> <li>soit il ne contient pas le champ obligatoire « ExecutableVersion » ;</li> <li>soit il contient un champ « ExecutableVersion »</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>soit ajouter un champ « ExecutableVersion » au griffon concerné ;</li> <li>soit rajouter une valeur dans le champ « ExecutableVersion » du griffon concerné.</li> </ul>

		sans valeur, alors que c'est obligatoire.	
	{ "ErrorDetail" : "Invalid date" }	Un champ de type « DATE » n'est pas écrit correctement, sous forme de date.	Vérifier que les champs « CreatedDate » du référentiel des griffons sont écrits sous forme de date.
<b>Import ou mise à jour d'un référentiel des scénarios de préservation</b>			
1	{ "ErrorDetail" : "Duplicate scenario : 'PSC-000001'" }	L'utilisateur a importé un référentiel contenant deux scénarios ayant le même identifiant.	Modifier l'identifiant de l'un des deux scénarios dont l'identifiant est égal à « PSC-000001 ».
2	{ "ErrorDetail" : "Invalid scenario for : 'null' : ['identifiant' : this field can not be empty]" }	Deux possibilités : <ul style="list-style-type: none"> <li>• soit un des scénarios ne contient pas le champ obligatoire « Identifiant » ;</li> <li>• soit un des scénarios contient un champ « Identifiant » sans valeur, alors que c'est obligatoire.</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>• soit ajouter un champ « Identifiant » au scénario concerné ;</li> <li>• soit rajouter une valeur dans le champ « Identifiant » du scénario concerné.</li> </ul>
3	{ "ErrorDetail" : "Invalid scenario for : 'PSC-000001' : ['name' : this field can not be empty]" }	Deux possibilités pour le scénario dont l'identifiant est « PSC-000001 » : <ul style="list-style-type: none"> <li>• soit il ne contient pas le champ obligatoire « Name » ;</li> <li>• soit il contient un champ « Name » sans valeur, alors que c'est obligatoire.</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>• soit ajouter un champ « Name » au scénario concerné ;</li> <li>• soit rajouter une valeur dans le champ « Name » du scénario concerné.</li> </ul>
4	{ "ErrorDetail" : "Invalid scenario for : 'PSC-000001' : ['griffinByFormat[0].timeOut' : this field must be strictly positive]" }	Deux possibilités pour le scénario dont l'identifiant est « PSC-000001 » : <ul style="list-style-type: none"> <li>• soit il ne contient pas le champ obligatoire « Timeout » ;</li> <li>• soit il contient un champ « Timeout » sans valeur, alors que c'est obligatoire.</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>• soit ajouter un champ « Timeout » au scénario concerné ;</li> <li>• soit rajouter une valeur dans le champ « Timeout » du scénario concerné.</li> </ul>
5	{ "ErrorDetail" : "Invalid scenario for : 'PSC-000001' : ['griffinByFormat[0].maxSize' : this field must be strictly positive]" }	Deux possibilités pour le scénario dont l'identifiant est « PSC-000001 » : <ul style="list-style-type: none"> <li>• soit il ne contient pas le champ obligatoire « MaxSize » ;</li> <li>• soit il contient un champ « MaxSize » sans valeur, alors que c'est</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>• soit ajouter un champ « MaxSize » au scénario concerné ;</li> <li>• soit rajouter une valeur dans le champ « MaxSize » du scénario concerné.</li> </ul>

		obligatoire.	
6	{ "ErrorDetail" : "Invalid scenario for : 'PSC-000001' : ['griffinByFormat[0].actionDetail[0].type' : this field can not be null]" }	Deux possibilités pour le scénario dont l'identifiant est « PSC-000001 » : <ul style="list-style-type: none"> <li>• soit il ne contient pas le champ obligatoire « type » ;</li> <li>• soit il contient un champ « type » sans valeur, alors que c'est obligatoire.</li> </ul>	Il faut : <ul style="list-style-type: none"> <li>• soit ajouter un champ « type » au scénario concerné ;</li> <li>• soit rajouter une valeur dans le champ « type » du scénario concerné.</li> </ul>
7	{ "ErrorDetail" : "Griffin 'TOTO' is not in database" }	Un des scénarios de préservation contient un griffon dont l'identifiant égal à « TOTO » n'existe pas.	Il faut créer ce griffon dont l'identifiant est « TOTO », puis importer le référentiel des scénarios de préservation.
8	{ "ErrorDetail" : "List: [TOTO] does not exist in the database." }	Un des scénarios de préservation contient un format dont le PUID égal à « TOTO » n'existe pas (dans le champ « FormatList).	Il faut corriger le scénario de préservation contenant le PUID en erreur (dans le champ FormatList).



## Annexe 3 : Paramétrages des scénarios de préservation

En fonction des griffons utilisés, un scénario pourra exécuter des actions de préservation en particulier et des arguments propres au griffon.

Ces arguments doivent être rédigés conformément au langage JSON et prennent la forme de chaînes de caractères enregistrées entre des doubles quotes dans un tableau de chaînes de caractères.

### A. ImageMagick

ImageMagick est un outil permettant de valider des formats, de générer des objets binaires et d'extraire des métadonnées techniques.

Il couvre pas moins de 200 types de formats de fichier<sup>49</sup>.

Type d'opération	Précisions sur l'action de préservation	Commentaires
ANALYSE	<pre>"ActionDetail": [   {     "Type": "ANALYSE"   } ]</pre>	
GENERATE	<pre>"ActionDetail": [   {     "type": "GENERATE",     "values": {       "Extension": "pdf"     }   } ]</pre>	Génération simple de binaires : on ne qualifie que l'extension de sortie.
	<pre>"ActionDetail": [   {     "type": "GENERATE",     "values": {       "Extension": "pdf",       "Args": [         "-quality",         "100",         "-units",         "PixelsPerInch",         "-density",         "72x72"       ]     }   } ]</pre>	<p>Génération plus complexe de binaires : en plus de l'extension de sortie, on peut ajouter des paramètres supplémentaires à apporter aux binaires à générer<sup>50</sup>.</p> <p>Point d'attention :</p> <ul style="list-style-type: none"> <li>la solution logicielle Vitam prend en charge les commandes d'appel aux objets binaires ;</li> <li>de fait, les paramètres faisant appel à un nom spécifique de fichier (ex : « -write ») ne sont pas supportés par la solution logicielle Vitam.</li> </ul>

49 Une liste des formats gérés par ImageMagick est disponible à la page suivante : <https://imagemagick.org/script/formats.php> (Lien consulté le : 14 février 2019).

50 Une liste des paramètres pris en charge se trouvent à la page suivante : <https://www.imagemagick.org/script/convert.php> (Lien consulté le : 19 février 2019).

## B. JHOVE

Jhove est un outil permettant uniquement de valider des formats.

Avec ce griffon, il n'est possible que de vérifier la validité des formats suivants : AIFF, ASCII, GIF, HTML, JPEG, JPEG 2000, PDF, TIFF, UTF-8, WAVE, XML, GZIP, PNG, WARC.

Type d'opération	Précisions sur l'action de préservation	Commentaires
ANALYSE	"ActionDetail": [ { "type": "ANALYSE", "values": null } ]	Il est <b>obligatoire</b> de passer un argument ou de mettre une valeur égale à « null ».
	"ActionDetail": [ { "type": "ANALYSE", "values": {"args": ["-strict"]} } } ]	

## C. Siegfried

Siegfried<sup>51</sup> est un outil permettant uniquement d'identifier des formats.

Type d'opération	Précisions sur l'action de préservation	Commentaires
IDENTIFY	"ActionDetail": [ { "type": "IDENTIFY" } ]	

## D. ODFvalidator

ODFvalidator<sup>52</sup> est un outil qui permet de valider des objets au format Open Document.

Type d'opération	Précisions sur l'action de préservation	Commentaires
ANALYSE	"ActionDetail": [ { "type": "ANALYSE" } ]	

## E. VeraPDF

VeraPDF est un outil qui permet de valider des objets au format PDF/A.

<sup>51</sup> L'outil n'est implémenté qu'à partir de la release 9 – version 2.3.0.

<sup>52</sup> L'outil n'est implémenté qu'à partir de la release 9 – version 2.3.0.

Type d'opération	Précisions sur l'action de préservation	Commentaires
ANALYSE	"ActionDetail": [ { "type": "ANALYSE" } ]	

## F. LibreOffice

LibreOffice<sup>53</sup> est un outil qui permet de convertir des fichiers bureautiques, notamment en PDF.

En fonction des formats générés, l'outil propose une liste d'arguments spécifiques<sup>54</sup>.

Type d'opération	Précisions sur l'action de préservation	Commentaires
GENERATE	"ActionDetail": [ { "type": "GENERATE", "values": { "Extension": "pdf" } } ]	Génération simple de binaires : on ne qualifie que l'extension de sortie.
	"ActionDetail": [ { "type": "GENERATE", "values": { "Extension": "pdf", "Args": ["pdf:SelectPdfVersion=1"] } } ]	Génération plus complexe de binaires : en plus de l'extension de sortie, on peut ajouter des paramètres supplémentaires à apporter aux binaires à générer <sup>55</sup> . L'exemple ci-contre contient par exemple un argument en vue de générer un fichier au format PDF/A.
	"ActionDetail": [ { "type": "GENERATE", "values": { "Extension": "csv", "Args": ["Text - txt - csv (StarCalc)", "43,34,UTF8"] } } ]	Génération plus complexe de binaires : en plus de l'extension de sortie, on peut ajouter des paramètres supplémentaires à apporter aux binaires à générer <sup>56</sup> . L'exemple ci-contre contient des arguments permettant de générer un fichier CSV encodé en UTF-8 et devant contenir des informations

53 L'outil n'est implémenté qu'à partir de la release 9 – version 2.3.0.

54 Une liste des paramètres pris en charge se trouvent à la page suivante : <https://github.com/LibreOffice/core/tree/master/filter/source/config/fragments/filters> (Lien consulté le : 21 février 2019).

55 Une liste des paramètres pris en charge se trouvent à la page suivante : <https://github.com/LibreOffice/core/tree/master/filter/source/config/fragments/filters> (Lien consulté le : 21 février 2019). Des informations spécifiques à la génération de fichiers PDF sont disponibles à la page suivante : [http://specs.openoffice.org/appwide/pdf\\_export/PDFExportDialog.odt](http://specs.openoffice.org/appwide/pdf_export/PDFExportDialog.odt) (Lien consulté le : 21 février 2019).

56 Une liste des paramètres pris en charge se trouvent à la page suivante : <https://github.com/LibreOffice/core/tree/master/filter/source/config/fragments/filters> (Lien consulté le : 21 février 2019). Des informations spécifiques à la génération de fichiers CSV sont disponibles à la page suivante : [https://wiki.openoffice.org/wiki/Documentation/DevGuide/Spreadsheets/Filter\\_Options](https://wiki.openoffice.org/wiki/Documentation/DevGuide/Spreadsheets/Filter_Options) (Lien consulté le : 21 février 2019).

	]	textuelles.
--	---	-------------

## G. Tesseract

Tesseract<sup>57</sup> est un outil qui permet d'extraire des métadonnées techniques et de générer un fichier texte destiné à contenir ces métadonnées extraites.

Type d'opération	Précisions sur l'action de préservation	Commentaires
GENERATE	"ActionDetail": [ { "type": "GENERATE", "values": { "Extension": "txt" } } ]	
	"ActionDetail": [ { "type": "GENERATE", "values": { "Extension": "txt", "Args": ["-psm", "3"] } } ]	

<sup>57</sup> Une liste des paramètres pris en charge se trouvent à la page suivante : <https://github.com/tesseract-ocr/tesseract/blob/master/doc/tesseract.1.asc> (Lien consulté le : 17 avril 2019).