vitam
archivage numérique

Vitam pour les développeurs : présentation fonctionnelle et technique

01 octobre 2020

Intervenant

Germain LEDROIT

Tech Manager

SOMMAIRE

01 Présentation fonctionnelle

02 Présentation technique

03 Présentation et utilisation des APIs

Présentation fonctionnelle

Vitam, c’est quoi ?

Vitam est un logiciel **Back Office** d’archivage numérique

- Pour l’archivage courant, intermédiaire ou bien même définitif
- Qui peut s’utiliser dans un contexte mutualisé (multi-tenant), pour gérer plusieurs entités utilisatrices
- Qui gère tout type d’archives numériques (SIP) si respect du **SEDA 2.1**

Vitam est libre et utilisable gratuitement : sous licence **Cecill V2.1**

Vitam se positionne au cœur du SI pour permettre d’assurer les fonctions d’archivage :

- En assurant l’indexation et la pérennisation
- En assurant la conservation de la valeur probante sur une **très grosse volumétrie**
- Il doit être interfacé (pour en faire un SAE) avec :
  - Des applications métier et des logiciels de services d’archives
  - Des infrastructures de **stockage**

Vitam : normes et standards

La solution Vitam respecte de nombreux normes et standards :

Normes Records Management

- ISO 15489
- NF Z 30-300…

Normes et réglementation Accès

- Règlement général sur la protection des données
- Directive réutilisation des informations du secteur public
- Code des relations entre le public et l'administration
- Loi informatique et Libertés (Loi CNIL)

Normes Archivage Numérique

- ISO 14721 (OAIS)
  - **O**pen **A**rchival **I**nformation **S**ystem
  - Modèle conceptuel destiné à la gestion, l'archivage et la préservation à long terme de documents numériques
- NF Z 42-013
  - Mesures techniques et organisationnelles pour l'enregistrement, le stockage, la restitution et la conservation de l'intégrité de documents numériques

Réglementation Classifiée de Défense

- Code de la Défense
- IGI 1300
  - Protection du secret de la défense nationale

Normes et réglementation métier

- Code du Patrimoine
- ISO 20614 / NF Z 44-022 / Seda
  - Modalité d'échanges pour l'archivage électronique (normalisation du SEDA)
- EAC-CPF, EAD, ISDF

Normes et réglementation Sécurité

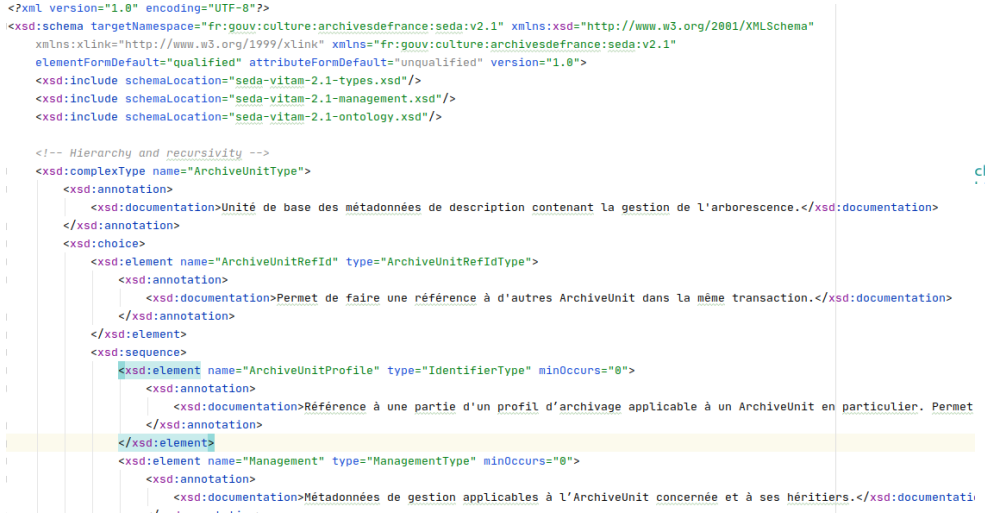
- eIDAS
  - Confiance dans les transactions électroniques
- NF Z 42-020
  - Description d'un coffre fort électronique au sein d'un SAE

C’est quoi le SEDA ?

Vitam parle le SEDA 2.1

- Standard d’Échange de Données pour l’Archivage
- Créé par les Archives de France et la DGME
- Basé sur l’OAIS, il modélise les transactions au sein d’un SAE
- Et assure l’interopérabilité entre les applications et le SAE
- Définit les Métadonnées descriptives

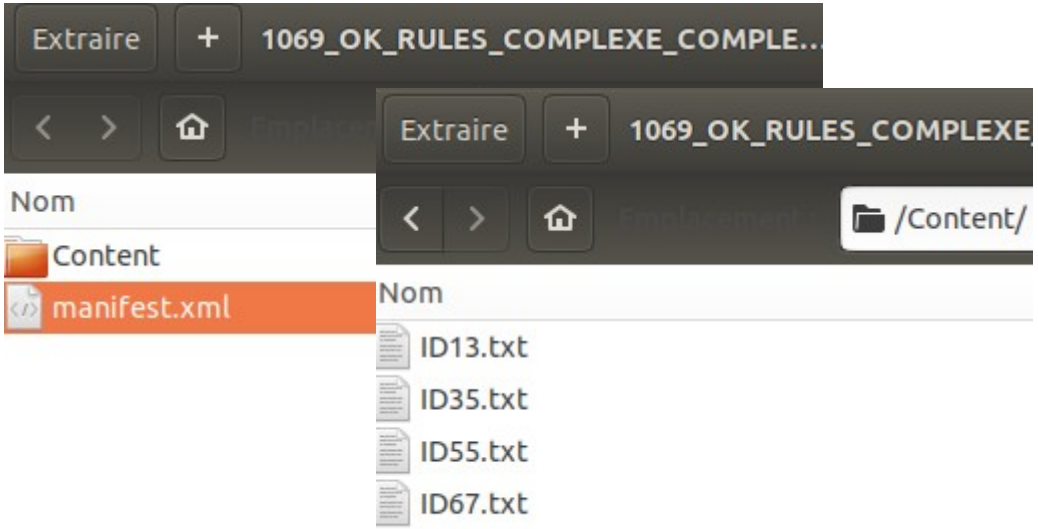
- Globalement, il s’agit d’une validation XSD extensible

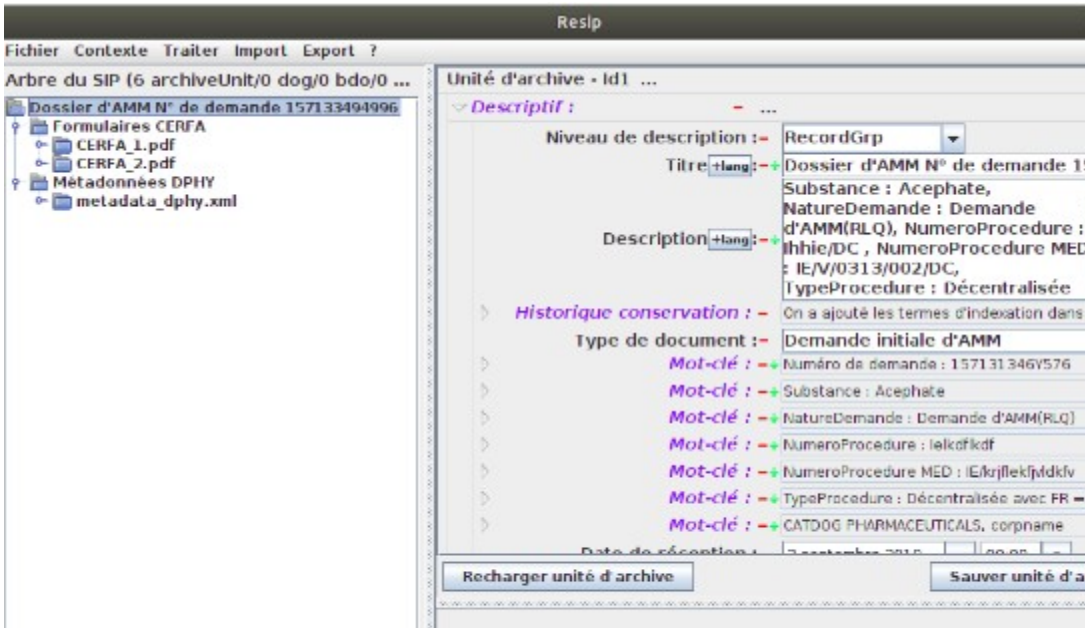


C’est quoi un SIP ?

Vitam accepte en entrée des SIP (Submission Information Package)

- Un SIP est un Zip ou un TAR (Tar, tar.gz, tarbz2 ou tar.gz2)
- Un SIP est composé de :
  - Un bordereau de transfert (manifest.xml) respectant le standard Seda 2.1. Ce bordereau contient des métadonnées représentant les binaires et des unités archivistiques
  - Un répertoire « Content » contenant les binaires à archiver.
- On peut créer un SIP avec un outil : ReSip, ou encore manuellement, si on est à l’aise avec le xml et le dictionnaire archivistique.

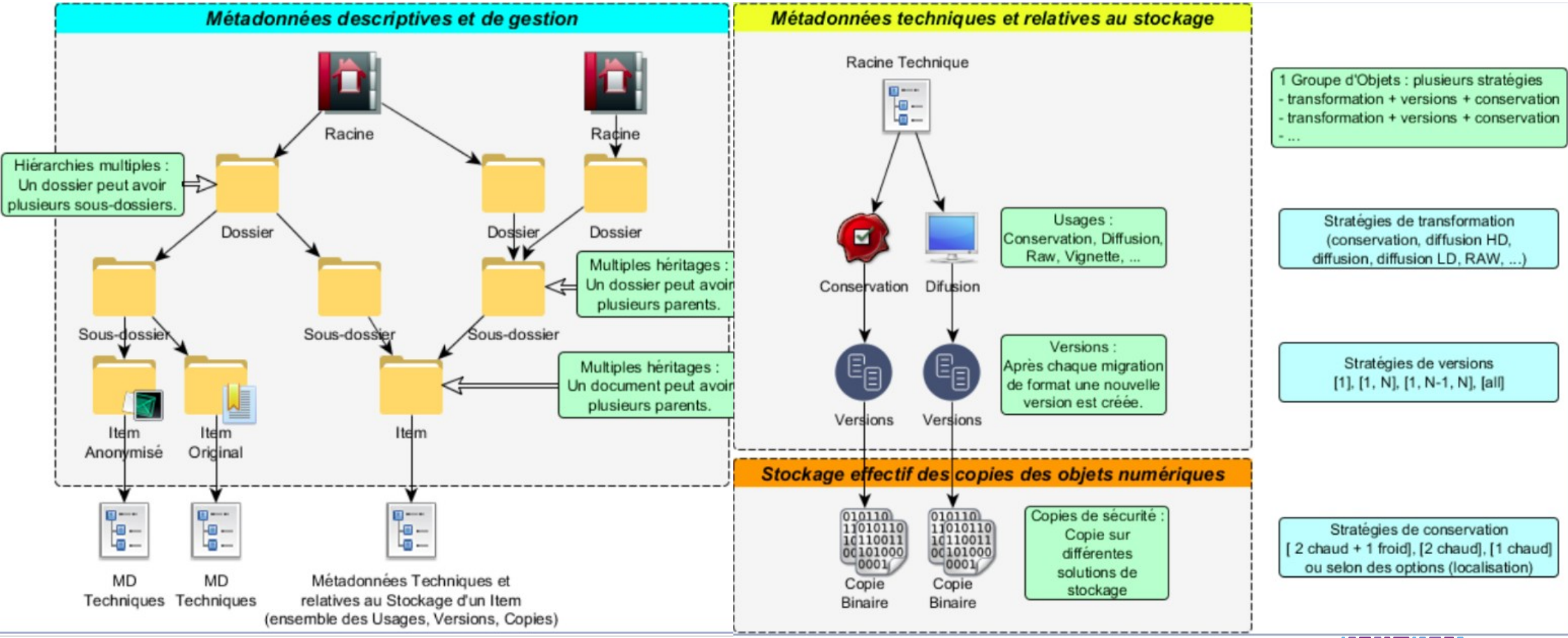




UA et GOT

UA : Unités Archivistiques (dossier avec Métadonnées)

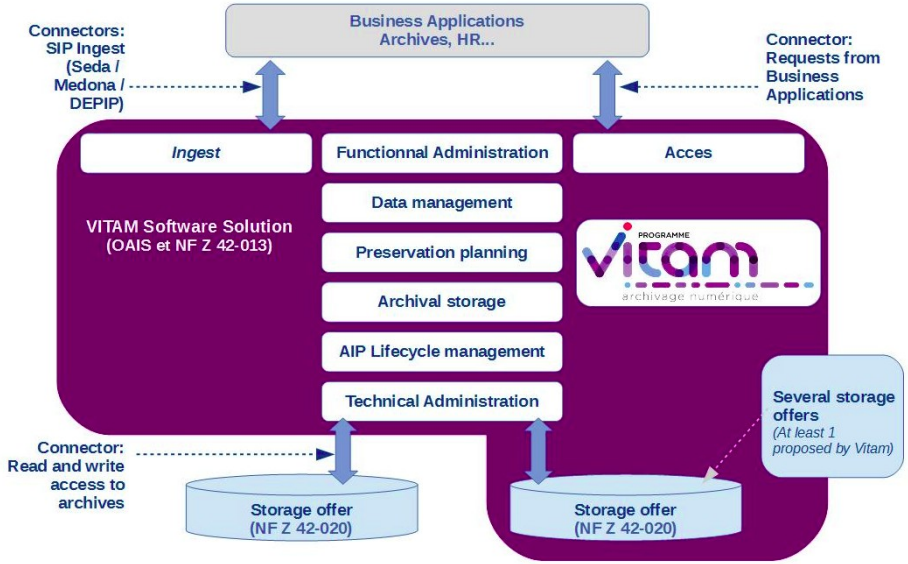
GOT : Groupe d'Objets Technique



Les fonctionnalités de Vitam

Vitam se base sur l'OAIS (Open Archival Information System), et est découpé en 6 grands domaines fonctionnels :

- Entrées : réception et traitement des entrées
- Stockage : stockage et conservation des archives
- Gestion des données : gestion des différentes Métadonnées
- Administration : supervision technico-fonctionnelle du système
- Préservation : gestion des formats, de la pérennisation, de conversion…
- Accès : recherche et consultation des données archivées



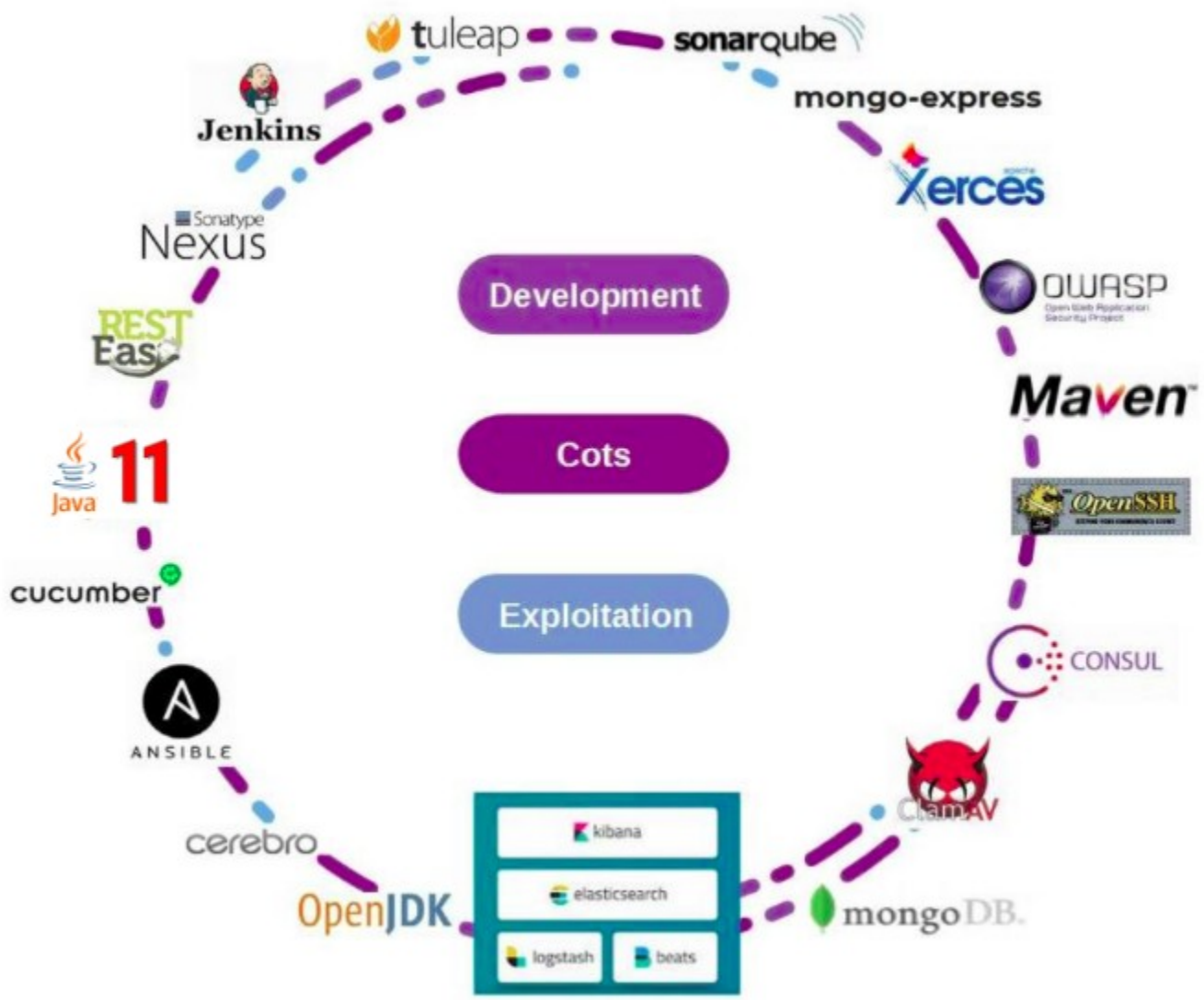
Journalisation et traçabilité

Étant un SAE (Système d'Archivage Électronique) à vocation probatoire, Vitam doit répondre à certaines normes, dont la NFZ 42-013 :

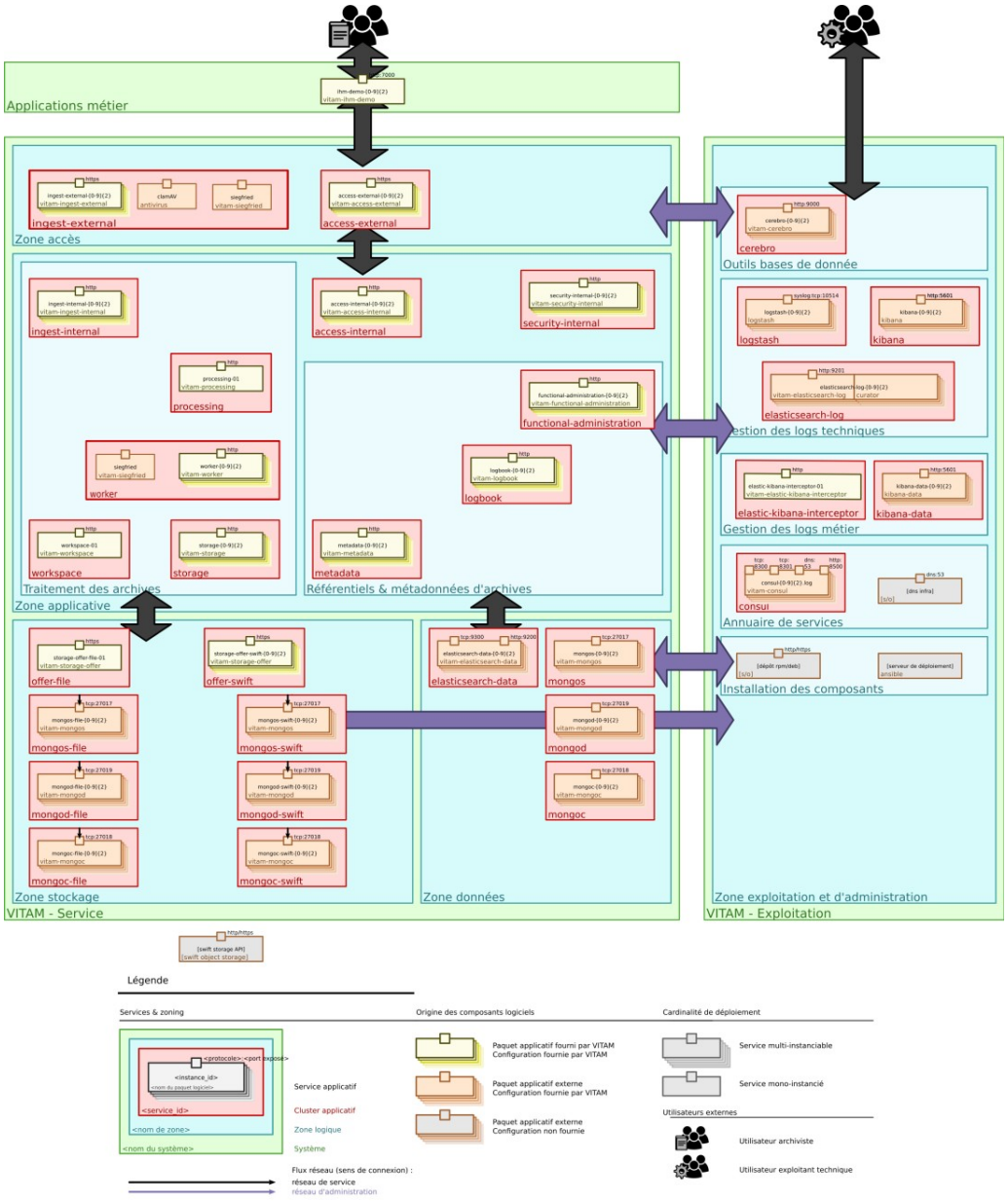
- Toute opération dans Vitam est tracée dans un journal des opérations :
  - Les opérations de support (gestion d'habilitations, de référentiels, audit, …)
  - Les opérations métier (versement, mise à jour d'archives, audit, élimination…)
- Les cycles de vie des AU et des GOT sont enregistrés dans un journal des cycles de vie
  - Ce journal doit contenir une entrée pour chaque création, modification, conversion, …
  - La journalisation du cycle de vie commence au versement initial de l'archive vers le SAE
- A intervalles réguliers, Vitam génère un fichier de sécurisation des différents journaux :
  - Chaîné (empreinte de sécurisation précédente est incluse dans la nouvelle)
  - Qui apporte de la sécurité cryptographique (arbre de Merkle)

Présentation technique

Socle Technique



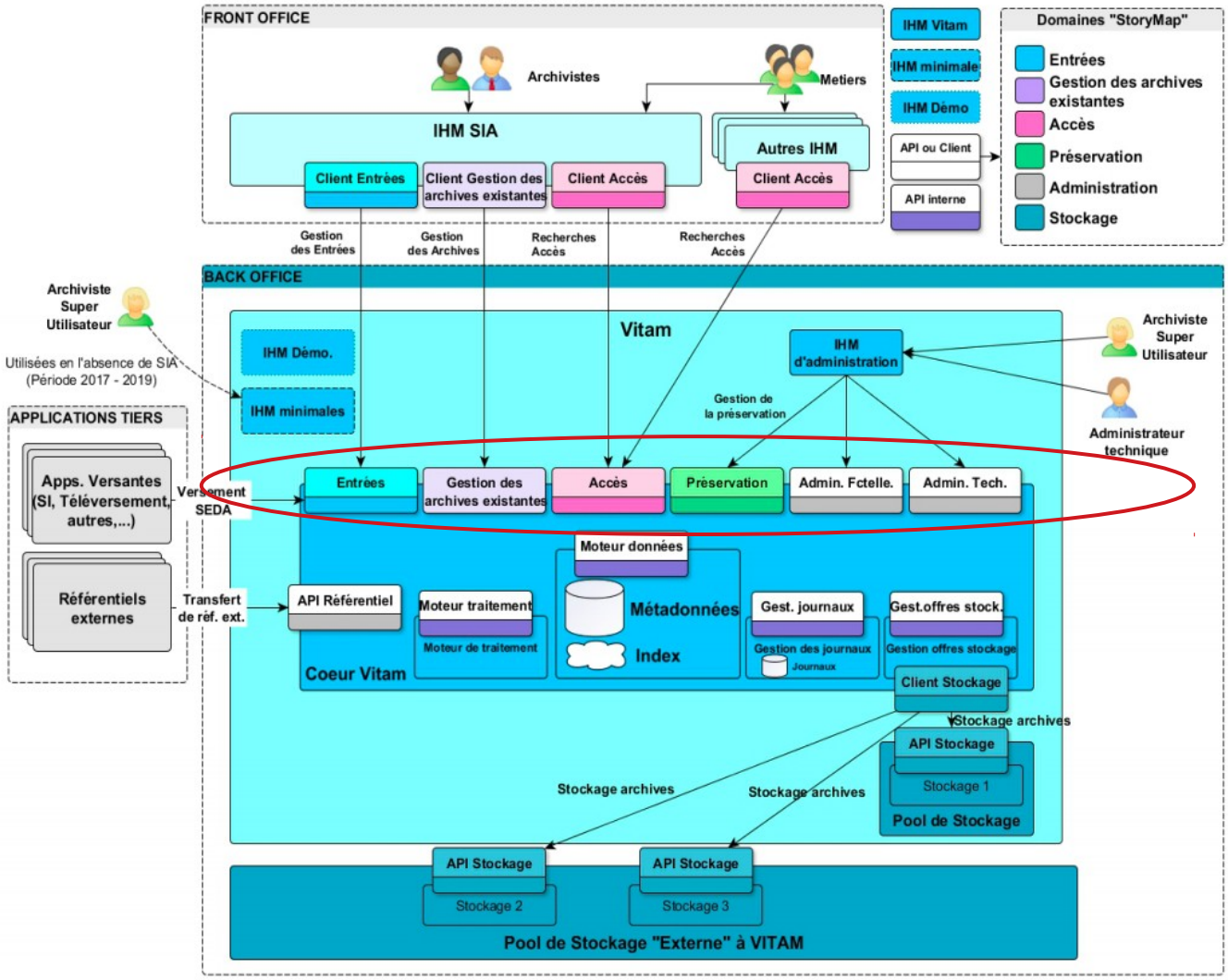
Architecture Technique



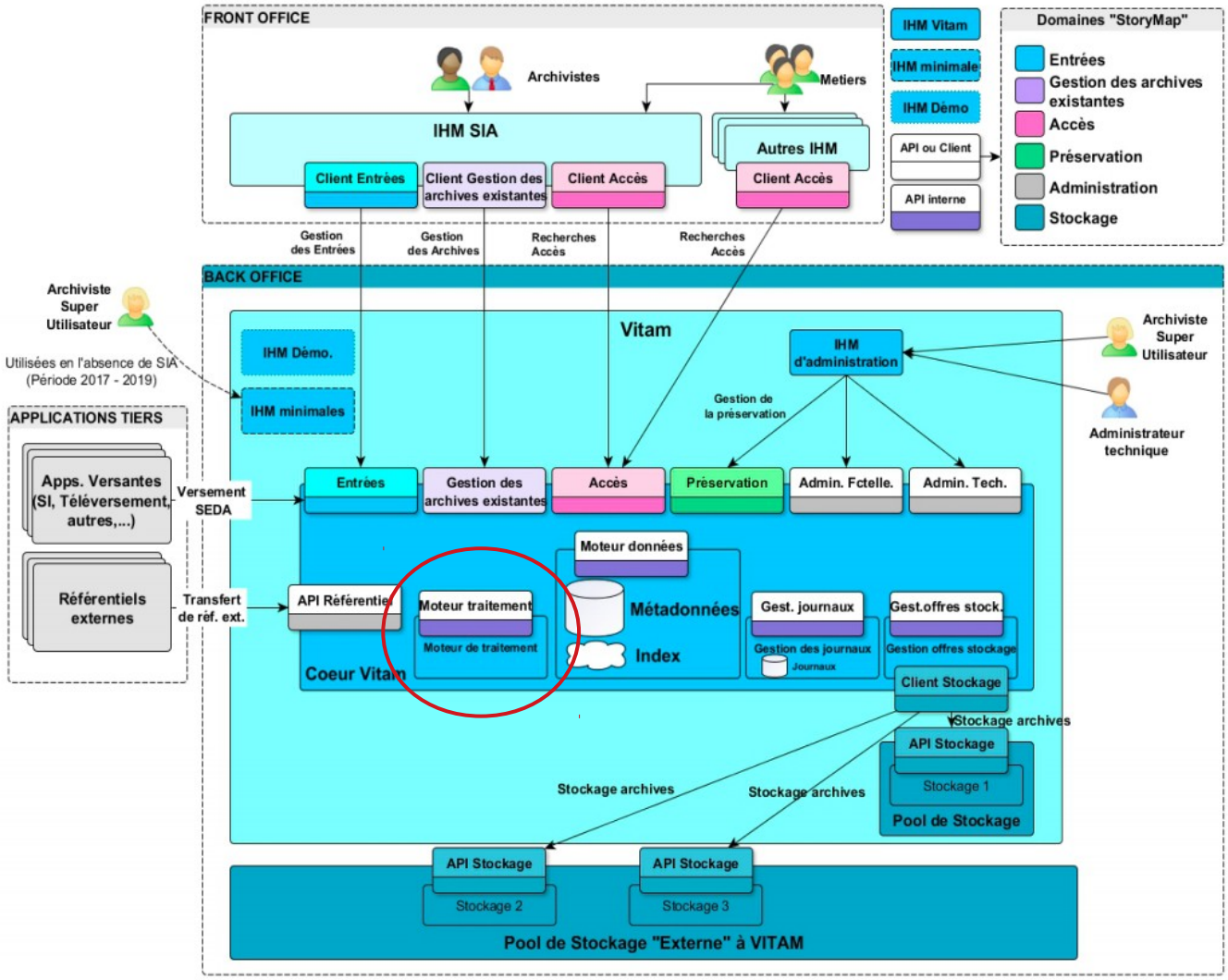
Principes d’architecture

- Architecture micro-service (+ de 20 modules répartis en différentes zones)
  - Modularité des composants : pour assurer la pérennité de la solution, et faciliter les migrations / changements
  - API Rest / Back End : communications via API Rest + utilisation d’un DSL (Domain Specification Language)
- Scalabilité / Résilience :
  - Duplication des modules + gestion fine des processus
  - PRA/PCA
- Différentes technologies d’offres de stockage acceptées : FS / S3 & Swift
- !! Grosse consommation mémoire (CPU et disque)
- !! Clusters MongoDB et ElasticSearch

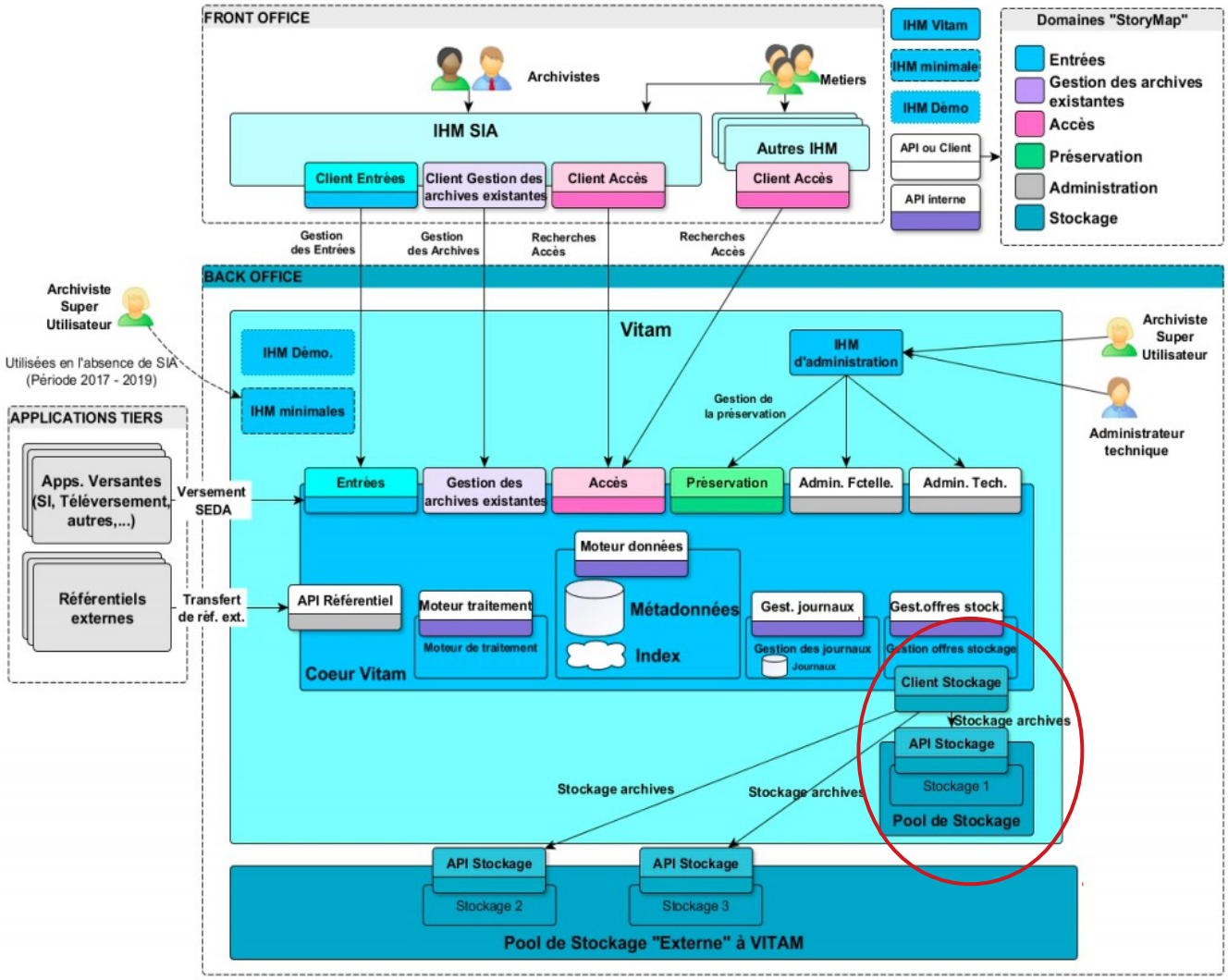
REST in peace



Le moteur d’exécution



Le stockage


FRONT OFFICE
Archivistes
Metiers
IHM SIA
Client Entrées
Client Gestion des archives existantes
Client Accès
Autres IHM
Client Accès
IHM Vitam
IHM minimale
IHM Démo
API ou Client
API interne
Domaines "StoryMap"
Entrées
Gestion des archives existantes
Accès
Préservation
Administration
Stockage
Gestion des Entrées
Gestion des Archives
Recherches Accès
Recherches Accès
BACK OFFICE
Archiviste Super Utilisateur
Utilisées en l'absence de SIA (Période 2017 - 2019)
APPLICATIONS TIERS
Apps. Versantes (SI, Téléversement, autres,...)
Versement SEDA
Référentiels externes
Transfert de réf. ext.
Vitam
IHM Démo.
IHM minimales
IHM d'administration
Archiviste Super Utilisateur
Administrateur technique
Gestion de la préservation
Entrées
Gestion des archives existantes
Accès
Préservation
Admin. Fctelle.
Admin. Tech.
Moteur données
API Référentiel
Moteur traitement
Moteur de traitement
Métadonnées
Index
Gest. journaux
Gestion des journaux
Journaux
Gest.offres stock.
Gestion offres stockage
Coeur Vitam
Client Stockage
Stockage archives
API Stockage
Stockage 1
Pool de Stockage
Stockage archives
Stockage archives
API Stockage
Stockage 2
API Stockage
Stockage 3
Pool de Stockage "Externe" à VITAM

Les Griffons

Pour assurer la pérennisation, Vitam intègre des composants « Griffons » :

- Analyse des fichiers/formats
- Ré-identification des fichiers/formats
- Conversion de fichiers/formats

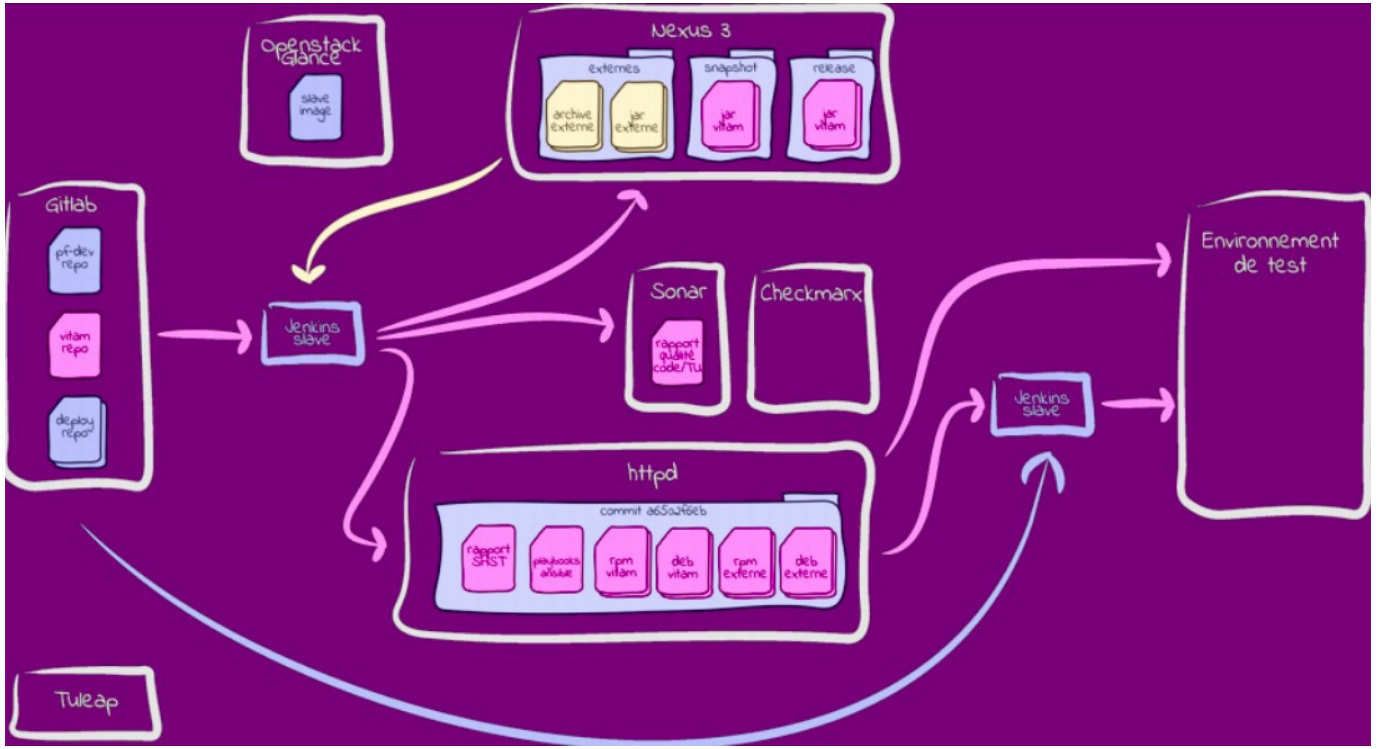
Quelques exemples de Griffons installés dans Vitam :

- Imagemagick : analyse, génération et extraction de données à partir d'une image
- Jhove : validation et analyse de formats
- Libreoffice : conversion de fichiers office en pdf
- Tesseract : Outil OCR pour extraire du texte dans une image
- ...

Déploiement

Le déploiement de Vitam se fait par des scripts ansible :

- Génération de packages DEB ou RPM
- Configuration dans des fichiers yml
- Possibilité d'automatiser les déploiements simplement
- Montées en version et opérations d'exploitation via des scripts ansible



Présentation et utilisation des APIs

Vitam, un Back-End exposant des APIs

Vitam est un back-end, qui expose des APIs regroupées en plusieurs catégories :

- Ingest
- Access
- Logbook
- Functional Administration

Basées sur un modèle REST classique, les ressources se basent sur un nom de collection. Des sous éléments peuvent compléter la ressource, ex : https://api.vitam.fr/access-external/v1/units/id

Les méthodes et les codes HTTP utilisées :

- Verbes : GET / POST / PUT / DELETE / HEAD / OPTIONS
- Codes HTTP : 2** / 4** / 5**
- Des headers sont parfois obligatoires (X-Application-Id, X-Tenant-Id....)
- Des requêtes asynchrones peuvent être exécutées et lancées (ordre d'une opération puis pooling jusqu'à finalisation de l'opération)
- Sécurité : certificats X509 avec authentification machine-machine + association avec contexte client (gestion accès, droits applicatifs).

Ingest

Base Url: https://api.vitam.gouv.fr/ingest-external/v1/[...]

* POST on /ingests (with consumedMediaTypes as an octet-stream): realises an ingest of the attached media (zip, tar, tar.gz or tar.bz2 file).

* POST on /ingests (with consumedMediaTypes as a json): realises an ingest of a file copied on a local disk (for big files)

* GET on /ingests/{id}/archivetransferreply/: get the ATR of an ingest

* GET on /ingests/{id}/manifests/: get the manifest of an ingest

Access

Base Url: https://api.vitam.gouv.fr/access-external/v1/[...]

* GET on /units search on units collection
* GET on /units/{id} get details of an archive unit
* GET on /units/{id}/objects/ (with “accept” as a json): get GOT information
* GET on /units/{id}/objects/ (with “accept” as an octet-stream): get object binary
* POST on /dipexport/: launch an operation of DIP generation
* GET on /dipexport/{id}/dip/: get the generated DIP
* POST on /transfers/: launch an operation of transfer (will generate a SIP)
* GET on /transfers/{id}/sip/: get the generated SIP
* POST on /elimination/analysis/: launch an analyse on elimination archives
* POST on /elimination/action/: launch an operation of elimination
* POST on /preservation/: launch an operation of preservation

[...] other endpoints

Logbook

Base Url: https://api.vitam.gouv.fr/access-external/v1/[...]

* GET on /logbookoperations search on operations
* GET on /logbookoperations/{id} get details of an operation
* GET on /logbookunitlifecycles/{id_lfc} get details of an unit Lifecycle
* GET on /logbookobjectslifecycles/{id_lfc} get details of an object Lifecycle

Functional Administration

Base Url: https://api.vitam.gouv.fr/admin-external/v1/[...]

* GET / POST on /accesscontracts to search/insert contracts

* GET / PUT on /accesscontracts/{idac} to get/update a contract details

* GET / POST on /rules to get/insert management rules

* GET on /rules/{idr} get details of a rule

* POST on /rulesfilecheck to check a rules insert

* POST on /audits to launch an audit operation

* GET on /operations: search amongst running or finished operation

* GET PUT DELETE HEAD on /operations/{idop}: get status of an operation, or update an operation (continue, stop, cancel, ...)

* GET on /workflows: get the list of workflows definition

* POST on /logbookoperations: to import into Vitam external logbooks

[....]

DSL

Le **DSL** (Domain Specific Language) est le langage dédié à **l'interrogation de la base de données et du moteur d'indexation** et offrant un niveau **d'abstraction** afin de pouvoir exprimer un grand nombre de possibilités de requêtes.

- **Request** : contient la structure du Body contenant la requête au format JSON. Le DSL permet d'exprimer un grand nombre de possibilités de requêtes. Dans le cadre des collections Metadata (Units et Objects), cette requête peut être organisée comme un ensemble de sous requêtes.
- **Response** : contient la structure du Body contenant le résultat au format JSON. Il contient les différentes informations demandées.

$**query** : la requête, composée de critères de sélection

$**roots** : les racines à partir desquels la recherche est lancée, uniquement pour les collections units et objects, dont les données sont organisées en mode arborescent

$**filter** : le tri / la limite en nombre de résultats retournés

$**projections** : un sous ensemble de champs devant être retournés

$**facets** : un tableau de requêtes d'aggrégation, uniquement pour la collections units

Quelques utilisations

Exemples de DSL :

- {"$roots": [], "$query": [{ "$match": { "Title": "Mon Titre" } }], "$filter": {}, "$projection": { "$fields": { "Title": 1, "Description": 1 }}}
- { "$roots": [], "$query": [{ "$or": [{"$match": { "Title": "Mon Titre" }}, {"$match": { "Title": "Mon autre Titre" }}] }], "$filter": {}, "$projection": {}, "$facets": [{ "$name": "facet_desclevel", "$terms": { "$field": "DescriptionLevel","$size": 5,"$order": "ASC" } }] }

Dans un cUrl :

- curl -v -X GET -k --key vitam-vitam_3.key --cert vitam-vitam_3.pem 'https://recetteur.env.programmevitam.fr/access-external/v1/units' -H 'X-Tenant-Id: 0' -H 'X-Access-Contract-Id: ContratTNR' -H 'Accept: application/json' -H 'Content-Type: application/json' --data-binary '{ "$roots": [], "$query": [{ "$and": [{"$exists": "Description"}, {"$match": { "Title": "Mon Titre" }}] }], "$filter": {}, "$projection": { "$fields": { "Title": 1, "Description": 1 } } }'

Quelques utilisations

Client java :

```
SelectMultiQuery searchQuery = new SelectMultiQuery();
CompareQuery queryTitle = QueryHelper.eq("Title", "Mon Titre");
searchQuery.addQueries(queryTitle);
AccessExternalClient accessClient = AccessExternalClientFactory.getInstance().getClient();
RequestResponse requestResponse = accessClient.selectUnits( new
VitamContext(tenantId).setAccessContract(contractId).setApplicationSessionId(applicationSessionId),searchQuery.getF
inalSelect());
[...]
```

Interfaçage SI

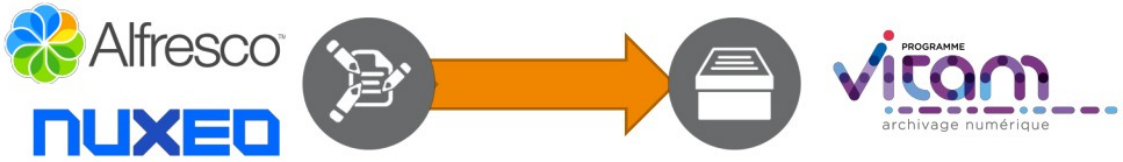
Pour **interfacer Vitam avec le SI existant :**

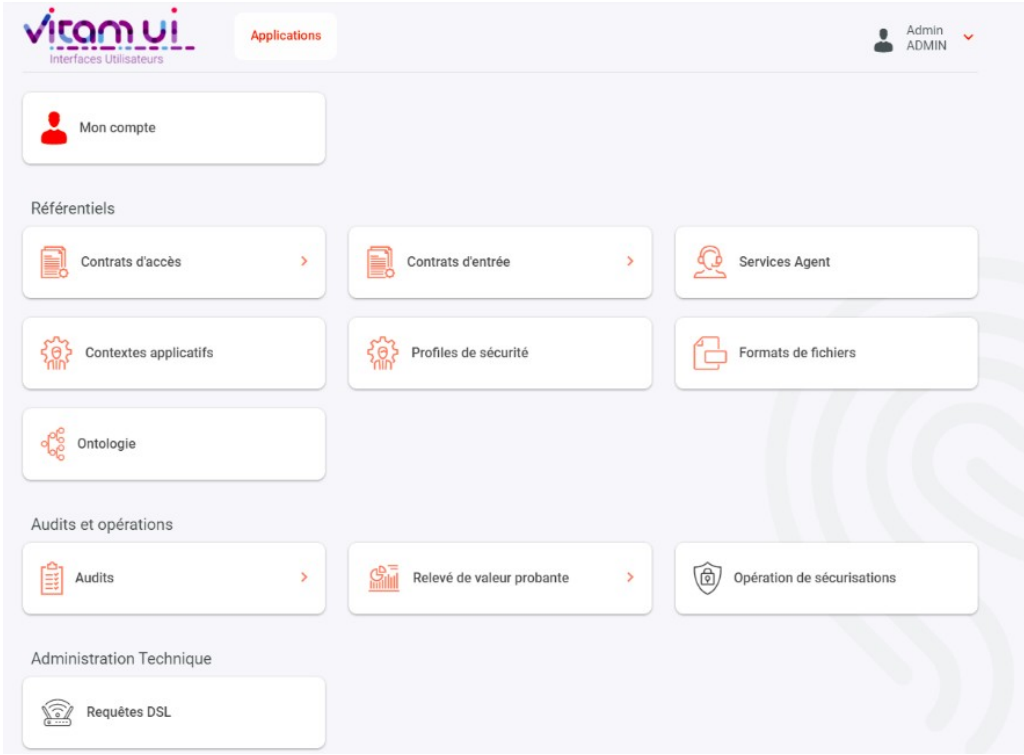
- Extraction des données existantes
- Connecteurs inter-applicatifs (ex : connecteur GED → VITAM)
- ETL
- Appel des API Vitam
- Intégration à d'autres composants du SI (anti-virus, monitoring, ...)
- IHM(s) à prévoir pour l'administration, le versement et l'accès

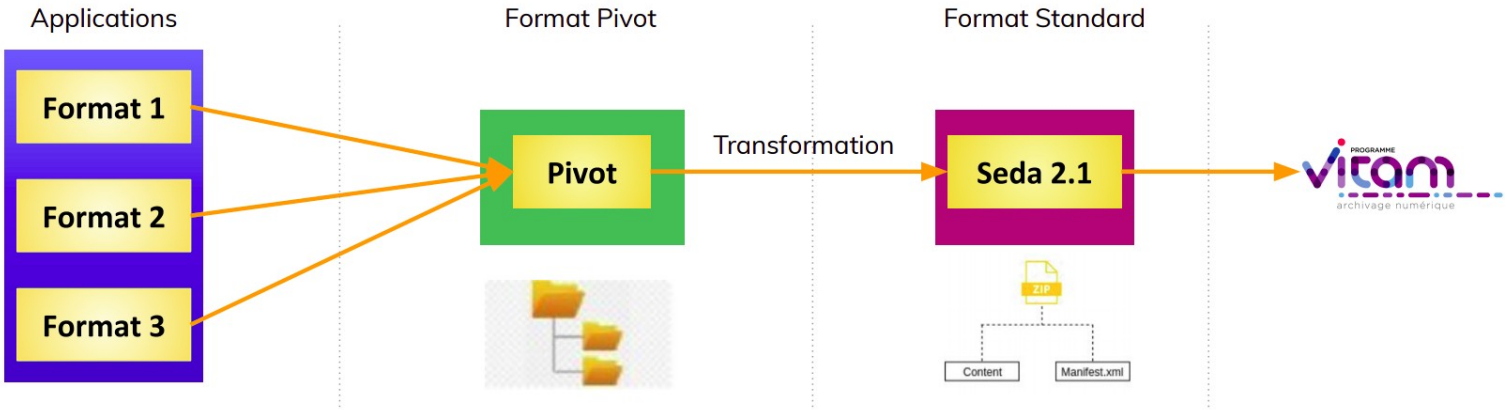
...

Interfaçage SI

Des **exemples :**







vitam
archivage numérique

Et si je veux tout savoir...

Code source :

- https://github.com/ProgrammeVitam

Documentation générale :

- http://www.programmevitam.fr/pages/documentation/

Documentation API :

- http://www.programmevitam.fr/ressources/DocCourante/raml/externe/

Programme Vitam

47 rue de la Chapelle, 75018 Paris – France
Tél. : +33 (0)1 86 69 60 03

www.programmevitam.fr
https://twitter.com/@ProgVitam
https://www.linkedin.com/grps/Programme-Vitam