



L'ontologie dans la solution logicielle Vitam

L'ontologie référence l'ensemble des vocabulaires (ou métadonnées) acceptés dans la solution logicielle Vitam. Pour chacun de ces vocabulaires, elle définit un intitulé et un type d'indexation particulier, ainsi que la ou les collections de la base de données de la solution logicielle Vitam qui l'utilise.

L'ontologie regroupe toutes les façons de nommer un même objet intellectuel. Elle permet d'éviter la multiplication des vocabulaires faisant référence à une même notion et de fiabiliser la manière dont ces vocabulaires et métadonnées sont indexés dans la solution logicielle Vitam pour optimiser les facultés de recherche.

PRÉSENTATION DE L'ONTOLOGIE

Périmètre de l'ontologie

L'ontologie est valable pour la plate-forme. Elle est dite trans-**tenant***.

Composition de l'ontologie

L'ontologie référence :

- les vocabulaires définis dans le SEDA, inclus par défaut,
- les vocabulaires propres à la solution logicielle Vitam, inclus par défaut,
- d'autres vocabulaires dits « externes » ajoutés pour répondre à un besoin de description particulier du service utilisateur non couvert dans les vocabulaires par défaut.

L'ontologie référence uniquement les vocabulaires pouvant contenir des valeurs. Cela signifie qu'un élément XML englobant un sous-élément XML (par exemple, les balises <Writer> ou <Management> du SEDA) n'est pas référencé.

L'ontologie définit pour chaque vocabulaire un type d'indexation qui correspond à la manière dont l'information doit être renseignée : attend-on une date, une chaîne de caractères, un mot-clé, un nombre entier, un nombre décimal, un booléen ?

Contrôles effectués à partir de l'ontologie

Lors de la création ou de la mise à jour d'un **profil d'unité archivistique***, la solution logicielle Vitam vérifie que les vocabulaires déclarés dans les profils existent dans l'ontologie.

La solution logicielle Vitam n'effectue pas de contrôle sur la conformité des vocabulaires par rapport à leur type d'indexation dans l'ontologie lors de la création de profils d'unité archivistique. Néanmoins, il est obligatoire que le type du vocabulaire défini dans le schéma de contrôle corresponde au type d'indexation du vocabulaire tel qu'il est défini dans l'ontologie.

Lors d'une entrée, la solution logicielle Vitam vérifie que :

- pour les unités archivistiques ne déclarant pas de profil d'unité archivistique, la valeur de leurs métadonnées est conforme à celle attendue dans l'ontologie (par exemple qu'une



Catégorie : contrôle des métadonnées, structuration des référentiels

métadonnée dont la valeur doit être une date est bien une date et pas un chiffre ou une chaîne de caractères), si ces métadonnées existent dans l'ontologie,

- pour les unités archivistiques qui déclarent un profil d'unité archivistique, la valeur de toutes leurs métadonnées est conforme à celle définie dans le schéma de contrôle associé.

Lors de la modification des métadonnées :

- pour les unités archivistiques ne déclarant pas de profil d'unité archivistique, si la métadonnée existe dans l'ontologie, la solution logicielle Vitam vérifie que la mise à jour demandée est conforme au type d'indexation du vocabulaire dans l'ontologie ;
- pour les unités archivistiques déclarant un profil d'unité archivistique, la solution logicielle Vitam vérifie que la mise à jour demandée reste conforme au profil d'unité archivistique et, de fait, à l'ontologie, aussi bien pour les métadonnées que les valeurs utilisées.

Utilisation de l'ontologie pour gérer les affichages dans l'interface graphique

L'ontologie contenant la traduction des différents vocabulaires supportés par la solution logicielle Vitam, il est possible, en accès :

- d'utiliser ce référentiel comme un fichier de propriétés pour récupérer les traductions, plutôt que de faire porter par l'interface graphique ces informations. Ainsi, cela évitera de constater des absences de traduction des vocabulaires externes récemment créés ;
- d'utiliser et d'afficher la traduction des vocabulaires dans les interfaces graphiques, rendue administrable dans ce référentiel, afin qu'un administrateur fonctionnel ait la possibilité de modifier les intitulés (ou traductions) de certains vocabulaires (par exemple, modifier « Description », traduction textuelle du bloc <Description> du SEDA, par « Présentation du contenu », terme issu de la norme ISAD/G).

CONFIGURATION DE L'ONTOLOGIE

Constitution de l'ontologie

La création d'une ontologie est un préalable à l'utilisation des vocabulaires dans la solution logicielle Vitam, c'est pourquoi lors de l'installation de la solution logicielle Vitam une ontologie est configurée par défaut. Cette ontologie contient l'ensemble des vocabulaires de type interne, c'est-à-dire les vocabulaires issus du SEDA et les vocabulaires propres à la solution.

Un administrateur fonctionnel n'a donc pas besoin, *a priori*, de créer une ontologie. Il s'agit d'un acte d'exploitation technique.

Néanmoins, dans certains cas, l'administrateur fonctionnel pourra être amené à vouloir enrichir l'ontologie par de nouveaux vocabulaires pour décrire des éléments propres aux archives qu'il souhaite prendre en charge qui ne figurent pas dans l'ontologie par défaut de la solution logicielle Vitam.

La création d'un nouveau vocabulaire n'est pas un acte anodin. Il est conseillé de prendre en considération les éléments suivants :

- est-ce qu'un vocabulaire déjà existant peut couvrir le même champ sémantique et signifiant ?
- pour quels usages a-t-on besoin d'ajouter ce nouveau vocabulaire ?
- est-il possible de rendre ce vocabulaire plus générique pour faciliter son utilisation dans un contexte plus large ? Par exemple, créer un vocabulaire « Somme » et non des vocabulaires



Catégorie : contrôle des métadonnées, structuration des référentiels

« SommeDepense », « SommeFraisDeplacement », « SommeDue », etc.

Transposition des outils existants dans une ontologie conforme aux attendus de la solution logicielle Vitam

Pour étudier son besoin de créer ou non de nouveaux vocabulaires dans l'ontologie, l'administrateur fonctionnel peut examiner les différents modèles de bordereaux de transfert utilisés dans son service pour étudier les notions métier décrites et leurs valeurs attendues (date, texte, nombre...) et déterminer ainsi son besoin en vocabulaires spécifiques non pris en charge dans le SEDA et leur type d'indexation.

Illustration pour un versement de dossier de personnel

Nom	Prénom	Date de naissance	Matricule	Corps	Grade	Nombre d'enfants	Date de départ en retraite	Décorations
-----	--------	-------------------	-----------	-------	-------	------------------	----------------------------	-------------

Les « Nom », « Prénom », « Date de naissance » et « Matricule » correspondent à des balises déjà existantes dans le SEDA. Le type d'indexation de ces balises est défini dans le SEDA.

La « Date de départ en retraite » pourra soit être exprimée à l'aide d'une des dates proposées par le SEDA (<EndDate> par exemple), soit par un vocabulaire spécifique <DateDepartRetraite> qui pourrait également être utilisé pour décrire des dossiers de pension. Le type d'indexation de cette <DateDepartRetraite> sera une date conforme au format date standardisé (exprimée par exemple sous la forme « 1794-02-01 » et non « 1^{er} février 1794 » ou « 13 pluviôse an II »).

Les « Corps », « Grade », « Nombre d'enfants » et « Décorations » ne correspondent à aucune notion présente dans le SEDA. Il conviendra donc de créer des vocabulaires spécifiques. Leur type d'indexation sera une chaîne de caractères, sauf pour le « Nombre d'enfants » où un nombre entier est attendu.

Illustration pour un versement de dossiers de marché

Type de marché	Numéro attribué aux marchés	Titulaire	Intitulé des marchés	Date de notification	Date de liquidation
----------------	-----------------------------	-----------	----------------------	----------------------	---------------------

Le « Type de marché » peut être traité par la balise <DocumentType> du SEDA.

Le « Numéro attribué aux marchés » peut être traité par les nombreux identifiants prévus par le SEDA.

Le titulaire sera une personne physique ou morale. La balise utilisée pour le décrire sera une balise englobante qui inclura des sous-éléments de type PersonOrEntityGroup ou BusinessGroup. Elle ne doit donc pas être déclarée en tant que telle dans l'ontologie.

L'intitulé du marché pourra correspondre à la balise <Title>.

Pour les « Date de notification » et « Date de liquidation », on pourra utiliser des balises du groupe DateGroup comme <TransactedDate> ou <EndDate> plutôt que de créer des vocabulaires spécifiques de type « DateDeLiquidation ».

Le type d'indexation des balises sera conforme à leur type d'indexation défini dans le SEDA.

Gestion des identifiants de l'ontologie

L'ontologie fonctionne en mode « **esclave*** ». Ce référentiel interne à la solution logicielle Vitam a



Catégorie : contrôle des métadonnées, structuration des référentiels

vocation à être une copie locale d'un référentiel administré dans le front-office des plates-formes d'archivage, de façon à permettre le partage des vocabulaires entre applications ou entre archives physiques et électroniques. C'est donc le front office qui fournit les identifiants, pour les vocabulaires externes uniquement.

Formalisation de l'ontologie

L'ontologie est importée dans la solution logicielle Vitam sous la forme d'un fichier au format JSON.

Un vocabulaire doit comporter les éléments suivants :

- **identifiant** unique dans l'ensemble du système (obligatoire) ;
- **nom** du vocabulaire (facultatif) ;
- **traduction** du vocabulaire, explicitant de manière intelligible le nom du vocabulaire (facultatif) ;
- **description** (facultatif) ;
- **type d'indexation** du vocabulaire, correspondant à un type attendu par le moteur de recherche Elastic Search (obligatoire) ;
- **origine** du vocabulaire, précisant la provenance du vocabulaire (facultatif). Sa valeur peut être égale à :
 - INTERNAL : pour les vocabulaires conformes au SEDA et les vocabulaires propres à la solution logicielle Vitam ;
 - EXTERNAL : pour les vocabulaires non gérés nativement par les deux précédents items et ajoutés pour répondre à un besoin particulier ;
- **collection(s)** de la base de données MongoDB qui utilise(nt) le vocabulaire en question (obligatoire).

Pour en savoir plus, consulter le document *Vitam. Modèle de données*.

MODIFICATION DE L'ONTOLOGIE

La modification de l'ontologie s'effectue par un ré-import de l'ontologie, en mode « annule et remplace ». C'est un acte d'administration lourd, car il peut nécessiter à la suite du ré-import une réindexation potentiellement longue. Cette réindexation est un acte d'administration technique.

La solution logicielle Vitam permet :

- d'ajouter un nouveau vocabulaire ;
- de modifier les informations associées à un vocabulaire (dont le type d'indexation) ;
- de supprimer un vocabulaire.

Cette action provoque la création d'une nouvelle version du référentiel.

La suppression d'un vocabulaire concerne exclusivement les vocabulaires d'origine externe, sauf en cas de mise à jour du modèle de données Vitam ou de publication d'une nouvelle version du SEDA.

La solution logicielle Vitam permet de paramétrer les possibilités de mise à jour de l'ontologie à l'initialisation de la plate-forme :



Catégorie : contrôle des métadonnées, structuration des référentiels

- elle peut interdire toute modification des vocabulaires internes, qui ont vocation à ne pas être modifiés ni surtout supprimés en production ;
- elle peut forcer la mise à jour des types d'indexation, sans tenir compte des règles prédéfinies.

Pour un vocabulaire externe et, si le paramétrage de l'ontologie le permet, pour un vocabulaire interne, les champs modifiables sont :

- le nom du vocabulaire :
 - tel qu'il est défini dans la nomenclature du SEDA ;
 - tel qu'il est retourné via la solution logicielle ;
- la traduction du vocabulaire ;
- la description ;
- le type d'indexation du vocabulaire, correspondant à un type attendu par le moteur de recherche Elastic Search ;
- l'origine du vocabulaire (les valeurs acceptées sont : INTERNAL, EXTERNAL) ;
- la collection de la base de données MongoDB qui utilise le vocabulaire en question (ce champ doit toujours contenir une référence à au moins une collection).

La mise à jour du référentiel peut échouer pour les motifs suivants :

- ajout d'un vocabulaire dont l'identifiant est déjà utilisé par un autre vocabulaire de l'ontologie ;
- ajout d'un vocabulaire dont l'identifiant ne correspond pas aux règles imposées par la solution logicielle Vitam ;
- incompatibilité entre le nouveau et l'ancien type d'indexation ;
- suppression d'un vocabulaire utilisé dans un profil d'unité archivistique.

Les opérations de modification des types d'indexation doivent obligatoirement être accompagnées d'un acte technique d'exploitation, afin que la nouvelle indexation soit effective et prise en compte par le moteur de recherche Elastic Search et que les données soient ré-indexées en base. La suppression d'un vocabulaire doit être accompagnée d'un acte d'exploitation technique visant à supprimer l'indexation du vocabulaire concerné dans le moteur de recherche Elastic Search, sans quoi le vocabulaire ne sera pas complètement supprimé de la solution logicielle Vitam. Pour plus d'information, consulter la documentation d'exploitation.

Les différentes versions de l'ontologie font l'objet d'une sauvegarde sur les offres de stockage utilisées par l'implémentation de la solution logicielle Vitam.

Pour aller plus loin, consulter le document *Vitam. Ontologie*.