

La structuration des informations dans la solution logicielle Vitam

Les archives nativement numériques sont des objets métier que les archivistes vont collecter, décrire, conserver et communiquer comme toutes les archives. Toutefois, leur support particulier entraîne certaines spécificités dans leur traitement archivistique par rapport aux archives traditionnelles sur support papier.

- 5 Deux normes principales concernent les archives numériques dans la sphère publique et sont à la base de l'organisation des informations dans la solution logicielle Vitam. La norme OAIS apporte un modèle conceptuel pour un système d'archivage électronique et définit la notion de « paquet d'information » à collecter, conserver et communiquer. La norme MEDONA modélise les transactions entre les différents acteurs des processus d'archivage (service d'archives, service de contrôle, service producteur, service versant, demandeur d'archives) et précise, dans sa déclinaison pour les archives publiques, le SEDA, la forme que prennent les informations échangées (constitution du paquet et formalisation de la description des métadonnées).

ORGANISATION INTERNE DE L'INFORMATION

- 15 La solution logicielle Vitam prend en charge, conserve et donne accès à :
- des unités archivistiques, avec leurs métadonnées descriptives et de gestion,
 - des groupes d'objets techniques qui regroupent les fichiers représentant les unités archivistiques,
 - des objets techniques avec leurs métadonnées techniques.

20 **Présentation d'une unité archivistique**

Une unité archivistique correspond à un niveau de classement intellectuel (ou unité de description). Elle peut décrire un fonds, un sous-fonds, une série organique, une sous-série organique, un dossier, un sous-dossier, un document, une page d'un registre...

- 25 Dans la solution logicielle Vitam, les unités archivistiques peuvent être de type **arbre de positionnement***, **plan de classement*** ou **standard***. Les unités archivistiques déclarent un niveau de description, au sens des normes de description archivistique (Series, RecordGroup, Item...), mais la solution logicielle n'effectue pas de contrôle de cohérence sur ces niveaux de description.

- 30 Une unité archivistique peut être rattachée à une ou plusieurs unités archivistiques (unités archivistiques parentes) et avoir une ou plusieurs unités archivistiques dépendantes (unités archivistiques filles), de manière à permettre une description archivistique hiérarchisée. Il est ainsi possible de naviguer dans une arborescence pour parcourir un fonds, une série, un dossier... Dans la solution logicielle Vitam, chaque point de cette arborescence est appelé **nœud***.

- 35 L'unité archivistique porte des métadonnées de description (au minimum, un titre et un niveau de description) et peut également comporter des métadonnées de gestion (**règles de gestion*** et/ou **profil d'unité archivistique*** applicables à l'unité).

La description peut être enrichie en fonction des besoins du service producteur. Le SEDA propose pour cela une liste des métadonnées descriptives et de gestion pouvant être utilisées et leur cardinalité (la métadonnée est-elle obligatoire ou non ? Est-elle répétable ?).

- 40 En cas de besoin, cette liste peut être complétée (cf. **ontologie***).

Les possibilités offertes par le SEDA permettent une description très riche des unités archivistiques : au-delà d'un titre et d'une date, il est également possible de faire porter à l'unité archivistique des identifiants archivistiques (cote) ou métier (numéro de dossier, matricule...) et des informations issues des systèmes de production métier (date de réception, d'envoi, de transaction, identité du rédacteur, du signataire et du destinataire...). Tout ou partie de ces métadonnées de description pourra par la suite faire l'objet d'un export pour alimenter des instruments de recherche, une base de données d'un portail... Il reviendra à l'archiviste de sélectionner la granularité de description pertinente et les métadonnées utiles.

Le traitement des archives électroniques implique une gestion fine à un niveau de granularité qui correspondrait au pièce à pièce dans le papier : en effet, le plus petit niveau d'unité archivistique pris en charge est le fichier. Par exemple, dans un dossier électronique de correspondance reçue par le Premier ministre, chaque courrier sera décrit séparément avec ses métadonnées. La notion d'article, qui combine unité matérielle et unité intellectuelle correspond ainsi dans le contexte électronique à l'objet nativement numérique pris en charge. Ceci permet une gestion très fine des accès puisqu'il est possible de cibler un objet très précis au sein d'un ensemble plus large pour le communiquer (par exemple, telle copie de concours, tel acte d'état civil, tel document dans un dossier).

Présentation des objets techniques

Les fichiers binaires sont appelés objets techniques. Ils sont typés en fonction de leur **version*** et de leur **usage***.

La version correspond à la génération technique d'un nouveau fichier, dans le cadre des opérations de **préservation***, pour assurer la lisibilité dans le temps.

La notion de version employée ici est différente de celle couramment utilisée pour désigner les versions successives d'un document dans le cadre d'un traitement administratif.

Pour une nouvelle version intellectuelle d'un document, on créera une nouvelle unité archivistique, puisque les documents ne seront pas identiques : contenu informationnel différent, date de dernière modification différente... Par exemple : nouvelle version de la documentation Vitam pour tenir compte des fonctionnalités développées.

Dans le groupe d'objets techniques, le nouveau fichier généré peut différer techniquement (par exemple, être un .pdf quand le fichier d'origine est un fichier .odt), mais son contenu informationnel est le même.

Les usages sont de quatre types pour les objets binaires : original numérique, version de diffusion, vignette, contenu textuel.

L'original numérique est le fichier sur lequel la solution logicielle Vitam fera reposer la valeur probante et dont elle s'engage à préserver l'accès dans le temps. C'est pourquoi, sauf exception, un groupe d'objets techniques doit toujours contenir un objet « original », qu'il s'agisse d'un original numérique (fichier) ou de la référence à un original physique conservé par le service.

Les objets techniques sont décrits par des métadonnées techniques (format, taille, date...). Le SEDA propose une liste des métadonnées techniques pouvant être utilisées et leur cardinalité (la métadonnée est-elle obligatoire ou non ? Est-elle répétable ?). La description technique peut être minimale ou enrichie en fonction des besoins du service producteur (cf. **ontologie***).

Présentation du groupe d'objets techniques

À chaque unité archivistique peut être associé un groupe d'objets techniques. Ce groupe d'objets regroupe tous les objets binaires qui représentent l'unité archivistique, selon leurs **usages*** et leurs **versions***, et peut référencer un objet physique (original physique) : un groupe d'objets peut ainsi contenir, par exemple, la référence à un registre, le fichier TIFF issu de la numérisation d'une des fiches du registre et le fichier en basse résolution utilisé pour la mise en ligne.

Le groupe d'objets techniques indique le nombre de copies des objets techniques réalisées par la solution logicielle en fonction de la stratégie de stockage appliquée.

90 Exemple de structuration des informations dans le cadre de la numérisation d'un registre matricule

Le registre sera décrit dans une unité archivistique associée à un groupe d'objets techniques qui référencera un original physique (le registre).

95 Chaque fiche matricule sera décrite dans une unité archivistique fille de l'unité archivistique décrivant le registre, associée un groupe d'objets techniques.

Le groupe d'objets techniques comprendra :

- un original physique (référence à la page du registre matricule),
- un original numérique (le fichier TIFF, éventuellement avec plusieurs versions au fur et à mesure des opérations de transformation de formats dans le cadre de la préservation),
- 100 • une copie de diffusion (le fichier JPEG, éventuellement avec plusieurs versions au fur et à mesure des opérations de transformation de formats dans le cadre de la préservation),
- une vignette (fichier basse résolution pour affichage d'un aperçu sur le site internet)
- un contenu textuel (fichier issu d'une opération d'océrisation de la fiche matricule).

CONSTITUTION DES PAQUETS D'INFORMATION

105 Le fonctionnement conceptuel d'un système d'archivage électronique repose sur la notion de paquets à soumettre, à archiver et à diffuser. On parle de « paquet » car l'objet à conserver (le fichier électronique, la base de données...) est obligatoirement accompagné des métadonnées descriptives et techniques qui vont permettre de le retrouver, de gérer son cycle de vie et de le préserver.

110 En entrée, la solution logicielle reçoit des SIP (Submission Information Package), constitués d'un bordereau de transfert et des fichiers à archiver.

Les fichiers à archiver sont contenus dans un répertoire informatique (la « boîte d'archives ») et sont accompagnés d'un bordereau de transfert.

115 Le bordereau de transfert, encore appelé manifeste, décrit les unités archivistiques, les groupes d'objets techniques et les objets techniques associés et pointe vers les fichiers à archiver. Ce bordereau de transfert correspond à la fois à une liste de colisage et au bordereau de versement des archives puisqu'il permet de faire le contrôle du nombre d'objets versés et porte également les descriptions des archives. Il peut faire l'objet d'un contrôle automatique de sa structure et des métadonnées renseignées (cf. **ontologie***, **profil d'archivage*** et **profil d'unité archivistique***). Il respecte un formalisme imposé par le SEDA.

120 La présence de fichiers numériques n'est pas obligatoire, ce qui permet de verser dans la solution logicielle Vitam des arborescences de description indépendamment des unités archivistiques avec leurs fichiers qui pourront y être **rattachées*** par la suite (cf. **arbre de positionnement*** et **plan de classement***).

- 125 Les SIP ne sont pas conservés tels quels dans la solution logicielle Vitam. Lors du processus d'entrée, chaque unité archivistique est enregistrée comme un AIP (Archival Information Package) qui référence son opération d'entrée dans la solution logicielle et sa ou ses unités archivistiques parentes. Ceci permet de reconstituer au besoin le SIP d'origine (tous les AIP ayant comme origine la même opération d'entrée) et l'arborescence des unités archivistiques, tout en offrant une grande
- 130 souplesse de gestion des archives au sein de la solution logicielle puisque les traitements archivistiques peuvent porter sur des unités archivistiques individuelles et non systématiquement sur le contenu d'une entrée.

- Lors des demandes de communication, la solution logicielle permet de constituer des DIP (Dissemination Information Package) composés d'un bordereau conforme au SEDA et des fichiers à
- 135 communiquer.
- Le bordereau décrit les unités archivistiques, les groupes d'objets techniques et les objets techniques associés et pointe vers les fichiers, réunis dans un répertoire informatique.
- Un DIP peut correspondre à une entrée, à une partie d'une entrée, à plusieurs entrées, à des unités archivistiques issues de plusieurs entrées.

140 **COEXISTENCE AVEC LES SYSTÈMES TRADITIONNELS DE DESCRIPTION ET DE GESTION DES ARCHIVES**

- Les services d'archives souhaiteront sans doute offrir à leurs publics une seule interface d'accès aux archives, quel qu'en soit le support. La coexistence des archives traditionnelles, numérisées et nativement numériques dans un même portail est tout à fait possible, dans la mesure où ces archives
- 145 partagent des caractéristiques communes, mais une attention particulière devra être portée sur certains points.

- Les archives nativement numériques sont décrites comme les archives traditionnelles de façon hiérarchisée, ce qui permet de naviguer dans une arborescence commune. Toutefois, le niveau de granularité de description des archives électroniques étant plus fin, cette spécificité est à prendre en
- 150 compte dans un système hybride pour gérer les descriptions de granularité très différente et leur restitution au lecteur : comment fournir des informations homogènes en évitant une surreprésentation des archives numériques ? Quel est le bon niveau de description des archives numériques à restituer : pièce, sous-dossier, dossier ?

- Exemple* : une messagerie électronique peut être décrite jusqu'au niveau de la pièce jointe à un
- 155 message. Dans un portail hybride, la description de la boîte de messagerie peut-elle suffire ?

- Les archives nativement numériques peuvent être accompagnées de métadonnées beaucoup plus riches que les métadonnées de description utilisées pour le papier : en plus du titre, des dates, de nombreuses informations issues des systèmes de production peuvent être conservées. L'archiviste devra s'interroger sur l'opportunité de publier ces métadonnées dans un portail ou un instrument de
- 160 recherche, tant pour conserver un équilibre avec les descriptions des archives traditionnelles que pour assurer une minimisation des éventuelles données à caractère personnel qui pourraient ainsi être exposées.

- La finesse de description des archives numériques permet de fournir une réponse simple à la notion de communication par extrait puisqu'il est désormais possible de cibler très précisément et finement
- 165 une unité archivistique (un courrier, un acte de naissance, un extrait signalétique des services...) au sein d'un dossier.