

KOST.07

Archivage des documents des systèmes de gestion des affaires (AUGev)

Résumé : modélisation des processus

Sommaire

1 Bases	1
1.1 Description du processus d'archivage actuel...	1
1.2 Délimitation des tâches à l'aide de l'OAIS ..	1
2 Modélisation des processus	2
2.1 Cycle de vie des documents numériques.....	2
2.2 Transférer des documents	3
2.2.1 Proposer des documents	3
2.2.2 Contrôler la proposition.....	4
2.2.3 Verser des documents.....	4
2.2.4 Réceptionner et vérifier des documents	4
2.2.5 Classer et décrire des documents	5
2.2.6 Éliminer des documents.....	5
2.3 Stocker des documents.....	5
2.4 Conserver des documents.....	6

Dans le cadre du programme de travail KOST.07, les Archives d'État de Bâle-Ville, de Zoug et de Saint-Gall ont mené à bien un projet de modélisation commun pour l'archivage des documents des systèmes de gestion des affaires (AUGev). Le présent document est un résumé des résultats de la phase conceptuelle, qui a essentiellement porté sur la modélisation des processus d'archivage. Un premier modèle de données a en outre été conçu. Pour ce qui concerne la documentation complète relative à cette phase de conception, nous renvoyons au document *Prozess- und Datenmodellierung* (modélisation des processus et des données).

Dans une seconde phase du projet, le modèle de données a été affiné, et les interfaces des systèmes ont été testées sur des prototypes et améliorées. Voir à ce sujet le *Résumé : modélisation des données et interfaces des systèmes*, ainsi que les autres documents concernant la phase II.

1 Bases

Le projet a eu deux points de départ : d'une part le processus actuel d'archivage des documents sur papier, d'autre part le modèle OAIS, référence conceptuelle de base pour l'archivage numérique et la délimitation des tâches.

1.1 Description du processus d'archivage actuel

Dans les processus reposant sur des documents papier, la phase de préarchivage et l'archivage propre-

ment dit peuvent être généralisés comme suit (l'ordre de succession adopté ci-dessous n'est pas contraignant) :

- Proposition de documents aux Archives, de la part d'un service versant
- Évaluation des documents par les Archives
- Préparation des documents au sein du service versant
- Versement des documents aux Archives
- Rédaction d'un bordereau de versement (registre des entrées aux Archives)
- Accusé de réception du versement délivré au service versant
- Contrôle de réception aux Archives
- Transfert dans d'autres boîtes d'archives et traitement conservatoire
- Traitement archivistique (classement et description)
- Réévaluation aux Archives (tri)
- Stockage
- Enregistrement de la cote d'emplacement dans l'instrument de recherche
- Information au service versant (élimination côté service versant)
- Traitement conservatoire ultérieur (maintenance des fonds)

1.2 Délimitation des tâches à l'aide de l'OAIS

Le modèle de référence OAIS décrit l'archivage numérique comme l'interaction entre des processus et des spécifications d'interfaces qui servent à enregistrer des documents numériques et à les rendre à nouveau disponibles.

Le projet porte sur l'archivage des documents des systèmes de gestion des affaires (système GEVER). Le champ du projet a été limité à des archives numériques, y compris la conservation des fonds, avec des interfaces vers un système GEVER, vers un instrument de recherche interne et vers l'utilisation. Le domaine du préarchivage, les modalités des instruments de recherche et l'utilisation des archives ne font donc pas partie du projet. Ces domaines ont cependant été intégrés dans la réflexion, notamment pour la spécification des interfaces.

Fondé sur le modèle OAIS, le projet s'est concentré sur les quatre processus centraux *Ingest* (transfert), *Archival storage* (enregistrement), *Data management* (gestion), *Access* (accès). Parallèlement, il a fallu spécifier les quatre interfaces *SIP* (interface avec le système GEVER), *AIP* (capsule de données), *DIP*

(format d'accès) et *Descriptive Info* (interface vers l'instrument de recherche), comme l'enseigne la figure 1.

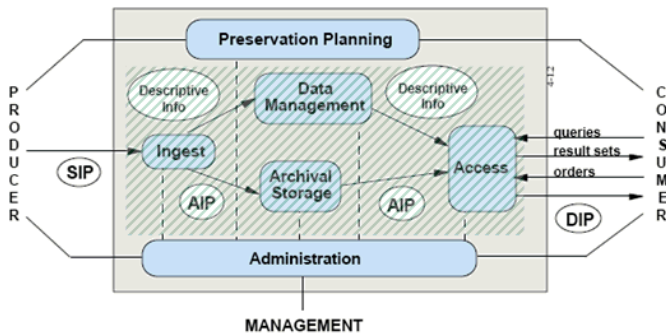


Figure 1 : Processus d'archivage centraux : Ingest, Archival Storage, Data Management, Access, représentés dans le contexte du modèle de référence OAIS¹

Une première étape de réalisation du modèle OAIS a porté sur l'attribution de fonctions aux couches logicielles concernées, comme l'explique le tableau I.

Couche logicielle	Fonction et description
Archival Management Software	L'instrument de recherche ou le catalogue permet un accès (structuré) aux données archivées
Digital Archiving Software	Réception, enregistrement, gestion et versement d'objets numériques
Storage Management Software	Gestion de l'emplacement d'enregistrement dans le système
Storage Hardware	Support physique d'enregistrement

Tableau 1 : Les principales couches logicielles

Le projet AUGev s'est concentré sur la couche « Digital Archiving Software ». Comme déjà dit, l'« Archival Management Software » n'est ici pertinent que pour la spécification des interfaces correspondantes. Plus bas, le « Storage Management Software » ainsi que le matériel (« Hardware ») correspondent à l'espace de stockage numérique. Celui-ci fait certes partie du domaine de tâches fixé, mais la spécification exacte et l'acquisition d'un espace de stockage numérique n'ont pour le moment été qu'esquissés dans le cadre du projet et feront l'objet du projet CECO arcun.

Les sous-processus selon OAIS peuvent être situés comme suit dans les couches logicielles:

- *Preservation Planning* et *Administration*: l'« Archival Management Software », également appelé « Archival Information System », peut contenir les processus accessoires OAIS *Preser-*

vation Planning et *Administration*. Nous avons renoncé à spécifier ici ces processus de gestion, car ils n'entrent pas dans le cadre fixé.

- *Ingest*, *Data Management*, *Archival Storage* et *Access* : il s'agit là des processus intrinsèques à la couche logicielle « Digital Archiving Software ». Un objet numérique (fichier) est réceptionné, enregistré et à nouveau livré.

2 Modélisation des processus

Dans le projet AUGev, le modèle OAIS a été utilisé pour modéliser des cas d'utilisation (« use cases »). Les différents cas d'utilisation ont été successivement examinés et spécifiés en détail. Le langage de modélisation UML (*Unified Modelling Language*) a été utilisé pour les structurer et les représenter. Ce langage permet une procédure de modélisation itérative : dans un premier cycle, on décrit l'interaction entre un utilisateur et le système modélisé, à l'aide de cas d'utilisation qui sont ensuite décomposés en cas d'utilisation supplémentaires, structurés en une succession d'activités et décrits en détails. Il se dégage alors une projection abstraite de tout le processus d'archivage vers différentes étapes et leurs résultats.

2.1 Cycle de vie des documents numériques

Le point de départ est le cycle de vie des documents numériques du point de vue des Archives, tel que l'esquisse le diagramme de cas d'utilisation de la figure 2.

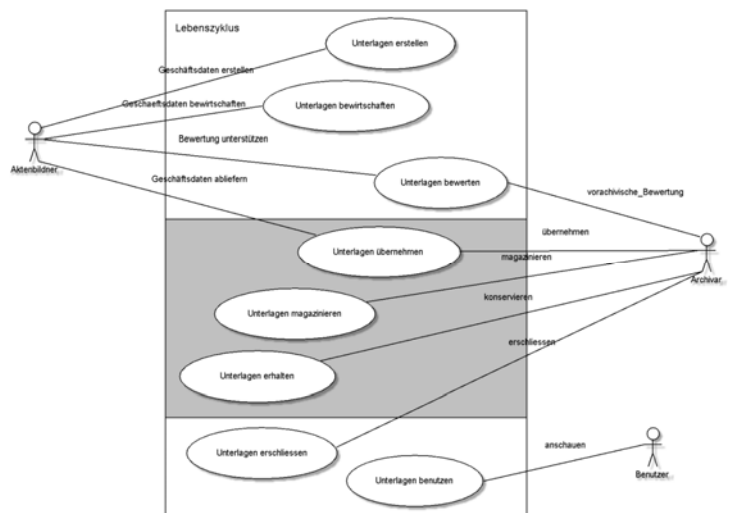


Figure 2: Cas d'utilisation Cycle de vie, dissociant les cas d'utilisation Provenance et les cas d'utilisation Utilisation

Le projet AUGev se concentre sur les cas d'utilisation d'archivage au sens étroit du terme (ci-dessus sur fond gris): *Transférer* (« übernehmen »), *Stocker* (« magazinieren ») et *Conserver* (« erhalten ») des documents (« Unterlagen »).

¹ Consultative Committee on Space Data Systems, Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, BLUE BOOK, January 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>

Évaluer (« bewerten ») des documents comprend l'évaluation et les conseils octroyés au service versant durant la phase de préarchivage, lorsque les documents se trouvent encore auprès du service versant. Traiter (« erschliessen ») des documents constitue un traitement archivistique postérieur, notamment au moyen d'indications fournies à partir de niveaux de description élevés (archives, fonds, versements). En contrepartie, l'étape Classer et décrire (« Ordnen und verzeichnen ») des documents, dans le processus Transférer (« übernehmen ») des documents, comprend l'intégration automatisée dans l'arborescence des archives et le transfert des métadonnées.

2.2 Transférer des documents

Le cas d'utilisation Transférer des documents (« Unterlagen übernehmen ») est décrit plus en détails dans la figure 3.

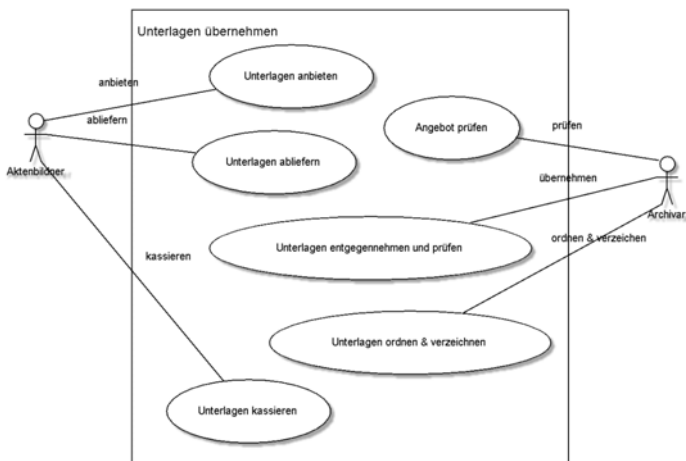


Figure 3: Diagramme du cas d'utilisation Transférer des documents, détails

Il en résulte le diagramme d'activités suivant (figure 4).

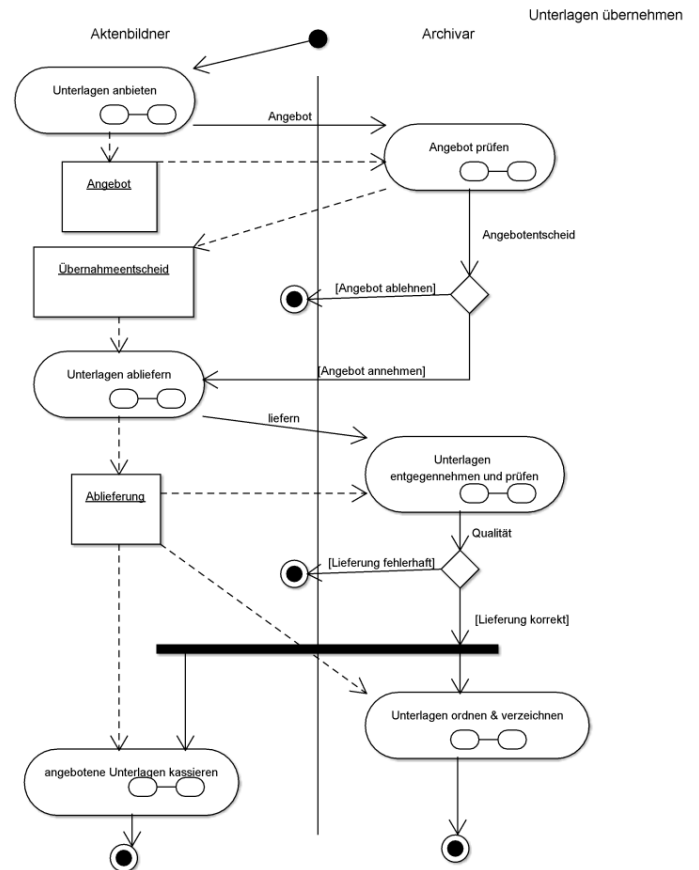


Figure 4: Diagramme d'activités Transférer des documents

Les différents cas d'utilisation sont spécifiés plus précisément ci-dessous.

2.2.1 Proposer des documents

Le service qui les produit propose aux Archives les documents destinés à la conservation définitive. Au moment de l'évaluation effectuée durant la phase de préarchivage, une convention identifie quels sont les documents destinés à être conservés définitivement, quelle est leur importance et à quelle fréquence ils sont versés. Une liste de tous les documents destinés à être versés est proposée ; elle contient des informations relatives au classement, aux métadonnées à y associer et au format de fichier. Cette proposition est adressée aux Archives de manière *ad hoc*. Les données primaires sont marquées dans le même temps comme « proposées » dans le système GEVER, et ainsi préparées pour pouvoir être envoyées ou pour pouvoir y accéder de l'extérieur.

En règle générale, la période de versement commence à la fin de la période de versement précédente et dure au maximum jusqu'à la date obtenue en retranchant de la date actuelle le délai de conservation. Le versement est constitué pour l'essentiel de dossiers clos ; une correction manuelle est également possible.

2.2.2 Contrôler une proposition de versement

Quand elles reçoivent une proposition de versement, les Archives contrôlent non seulement la quantité et, par échantillonnage, la qualité des documents proposés, mais comparent également la proposition à l'aune du référentiel de gestion. Il s'agit d'une part d'une vérification automatisée selon des critères formels, d'autre part d'un contrôle sur le fond. Les Archives, lors de ce dernier contrôle, effectuent des échantillonnages pour voir si les documents proposés correspondent au référentiel de gestion et aux intérêts archivistiques. Pour ce qui concerne les pièces d'une série documentaire, la taille et la pertinence statistique de l'échantillonnage sont également examinées. Nous partons de l'idée, dans le cadre de notre projet, qu'une évaluation prévisionnelle a déjà eu lieu. L'examen sur le fond, qui ne part donc pas de rien, consiste plutôt à vérifier l'application des critères d'évaluation, à les corriger, le cas échéant, et à pratiquer un éventuel échantillonnage.

Il se peut que les Archives refusent la proposition pour des raisons de forme ou de fond et qu'elles élaborent une nouvelle convention de versement avec le service versant. Sinon, une décision de transfert est communiquée sous forme d'une liste des documents à verser. Cette liste peut reprendre l'intégralité des documents proposés, mais aussi n'en englober qu'une partie.

2.2.3 Verser les documents

Le service versant prépare sous une forme adaptée au versement les documents (données primaires et métadonnées) cités dans la décision de transfert. Pour cela, il vérifie la forme et le contenu de la décision et l'assigne à une proposition ouverte. Les documents sont exportés en format versement depuis le système GEVER du service versant. Des métadonnées de procédure et des métadonnées techniques sont ajoutées en vue du versement. Le statut des documents versés doit en outre être ajouté dans le système GEVER.

Quand les documents réclamés sont prêts à être versés, sous la forme qui convient, ils sont soit retirés par les Archives (*pull*: espace FTP), soit envoyés aux Archives (*push*: Filetransfer ou autre).

2.2.4 Réceptionner et vérifier des documents

Au moment du versement, les Archives commencent par vérifier que tous les documents mentionnés dans le bordereau de versement y figurent réellement. Si celui-ci est incomplet, il est retourné, et le processus est stoppé. L'exactitude syntaxique des métadonnées est également vérifiée (vérification de la syntaxe XML et validation de schémas).

Si le versement est correct, on procède à la validation du format des données primaires versées, c'est-à-dire des fichiers contenant les documents à proprement parler. Les informations nécessaires à ce processus

(logiciel de reconnaissance, outil de validation, pérennité) figurent dans le catalogue des formats. Les résultats sont indiqués dans le journal des entrées.

Les informations relatives aux formats des documents contenus dans le versement donnent statistiquement une idée de la qualité du versement. Il ne serait ni judicieux, ni pratique de refuser un versement pour le seul motif que quelques documents n'ont pas un format de fichier valide. La décision d'accepter ou de refuser un versement doit plutôt se baser sur la qualité de l'ensemble, c'est-à-dire sur la proportion de documents défectueux. Quoi qu'il en soit, un versement incomplet ou insuffisant du point de vue formel doit conduire à refuser le versement.

Les Archives informent le service versant si le versement est correct. Cette quittance spécifie pour chacun des documents proposés s'il est archivé ou non par les Archives. Les documents doivent ensuite être éliminés dans le système d'origine, c'est-à-dire physiquement effacés.

Si le versement est correct, il est repris en vue d'un traitement ultérieur (archivage numérique à long terme).

Cette procédure implique que le cas d'utilisation *Réceptionner et vérifier des documents* («*Unterlagen entgegennehmen und prüfen*») comporte d'autres cas d'utilisation. Ceux-ci sont spécifiés dans le diagramme de cas d'utilisation qui suit (figure 5) et sont détaillés ci-après.

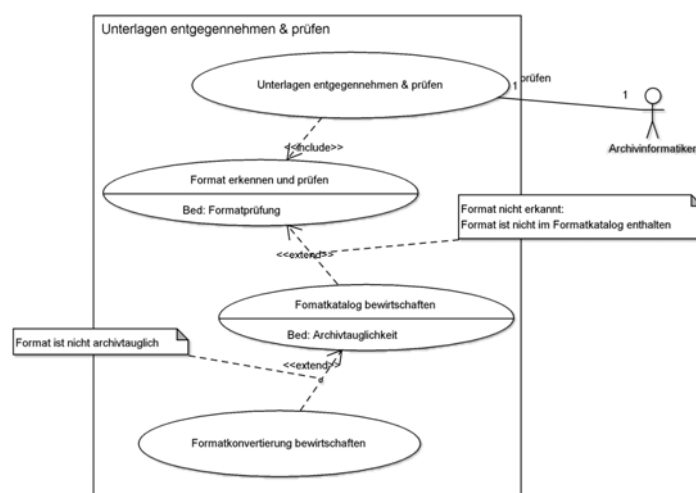


Figure 5 : Diagramme du cas d'utilisation Réceptionner et vérifier des documents

2.2.4.1 Reconnaître et valider le format

La validation du format est spécifiée comme un cas d'utilisation en soi, qui provient du cas d'utilisation *Réceptionner et vérifier des documents*. Le format d'un fichier est reconnu à l'aide des informations figurant dans le catalogue des formats, et le fichier est validé au moyen d'un outil de comparaison avec

la spécification du format. Si la validation échoue, une clarification manuelle est nécessaire. Les informations relatives à la pérennité des fichiers sont ensuite tirées du catalogue des formats. Les résultats de ces contrôles sont consignés dans le journal des entrées.

2.2.4.2 Tenir à jour le catalogue des formats

Si un fichier est versé dans un format qui n'a pas encore été répertorié, ce format doit être ajouté dans le catalogue des formats. Il faut pour cela tout d'abord déterminer de quel format il s'agit, généralement sur la base des informations déjà données par le service versant.

On consigne ensuite dans le catalogue des formats, pour chaque format et pour chaque version d'un format, les caractéristiques pour le reconnaître, le programme qui permet de valider un fichier en le comparant aux spécifications de ce format et le programme à utiliser pour restituer ce format (dans la version *ad hoc*). On évalue également si le format est pérenne ou non, en se fondant sur le *Catalogue des formats de données d'archivage Cfa*. Cette indication est, elle aussi, consignée dans le catalogue des formats. Si le format n'est pas pérenne du point de vue de l'archivage, le processus *Gérer la conversion de format* est appliqué.

Pour un format pérenne et déjà existant, le visionneur est évalué et éventuellement mis à jour. La pérennité est réévaluée.

2.2.4.3 Gérer la conversion de format

Lorsqu'un format n'est pas pérenne, un chemin de conversion est déterminé et intégré au catalogue des formats. On détermine, dans cette perspective, un format cible pérenne pour un format source périmé et on choisit un logiciel de conversion. Il convient d'intégrer au catalogue des formats tout format cible qui n'y figure pas. (Remarque : il se peut également qu'il n'existe aucun format cible pérenne, dans lequel il soit possible de convertir directement le format source. Dans ce cas, il faut contourner le problème en choisissant un format intermédiaire.)

2.2.5 Classer et décrire des documents

Une fois le fond et la forme examinés, on constitue un *fonds* avec les documents d'un versement. (Le *fonds* désigne ici l'ensemble des *versements* qui proviennent d'un service, d'un département ou de toute autre entité administrative ; se référer au *glossaire AUGev*.) La constitution d'un fonds, regroupement logique des versements par *service* et par *système de classement des dossiers courants*, est fondamentale pour l'enregistrement du versement dans l'instrument de recherche, et canalise le transfert des métadonnées GEVER dans l'instrument de recherche.

On recherche (automatiquement) dans les instruments de recherche pour tout versement le fonds correspon-

nant. Les Archives ont développé deux méthodes pour intégrer un versement à un fonds ; toutes deux doivent être supportées.

La première méthode consiste à intégrer le versement en tant que tel, avec sa propre structure, au fonds du service versant. Un fonds est ainsi constitué d'un ensemble de versements structurés en eux-mêmes (ex. : les Archives fédérales suisses).

La deuxième méthode consiste à intégrer plusieurs versements dans un fonds, c'est-à-dire à créer une application entre le système de classement du versement et l'arborescence des Archives, pour autant que leur concordance soit suffisante, et à conserver les dossiers du nouveau versement à l'emplacement correspondant. Si le système de classement du service versant change radicalement, un nouveau fonds est ouvert dans les Archives (par exemple Archives d'État de Bâle-Ville).

S'il n'y a encore aucune mention dans l'arborescence des Archives, il convient d'ouvrir un nouveau fonds dans l'arborescence.

Les informations Fonds, Versement, Série, Dossier numéroté en continu et Document numéroté en continu génèrent la cote de chaque dossier ou document versé :

Archives : [Service] : [Fonds] : [Versement] : [Série] : N° dossier : N° document

Cette cote est introduite dans le versement. Elle est le GUID (*Globally Unique Identifier*) du dossier et du document et en permet l'accès en cas de besoin. Toutes les métadonnées, y compris la cote ainsi ajoutée, peuvent ensuite être enregistrées dans l'instrument de recherche en une seule opération. Cela se fait de manière automatisée, à l'aide de l'assistant d'importation correspondant.

2.2.6 Éliminer des documents

Une fois proposés, tous les documents doivent être marqués en vue de leur élimination par le service versant. Ce dernier reçoit ainsi, pour tous les documents qui ont été soumis, une attestation sous une forme définie, qu'il est donc possible d'automatiser. Cette attestation précise quels documents ont été archivés et lesquels ne l'ont pas été.

2.3 Stocker des documents

Quand la conformité formelle des données primaires et des métadonnées réceptionnées a été vérifiée, qu'elles ont été intégrées dans un fonds et pourvues d'une cote, elles peuvent être stockées sur une plateforme d'archivage à long terme des données électroniques. Cette plateforme, appelée ci-après « espace de stockage numérique », doit remplir un certain nombre d'exigences (critères fonctionnels et non fonctionnels). Se référer sur ce point au projet *arcun*.

Un paquet d'information est formé des données primaires et des métadonnées y relatives. Ces métadonnées sont :

- les métadonnées GEVER livrées
- des métadonnées techniques, générées lors de la vérification formelle, au cours du processus *Transférer des documents*. Ces métadonnées sont transférées du journal des entrées dans le paquet d'information.
- l'historique des migrations du paquet d'information
- la *pack-list*, qui rend le contenu du paquet d'information accessible pour nous
- la cote

Le paquet d'information est identifié au moyen de la cote générée au cours du cas d'utilisation *Classer et décrire des documents* et est implémentée avec cette cote dans l'espace de stockage numérique en tant que GUID.

Il se peut que le paquet d'information contienne des fichiers dont le format n'est pas pérenne. C'est la raison pour laquelle ont fini par faire appel au cas d'utilisation *Migrer des documents*, spécifié dans le cadre du cas d'utilisation *Conserver des documents*.

2.4 Conserver des documents

Si, dans des délais raisonnables et pour des raisons techniques, il n'est plus possible de lire le contenu et/ou de restituer la forme de présentation d'origine de données, il faut les migrer dans un nouveau format de fichier (Condition préalable : l'infrastructure qui enregistre les données est fonctionnelle en tout temps, c'est-à-dire que le flux des bits reste lisible, le but étant de permettre de lire le contenu des données et de restituer la forme de présentation d'origine de ces dernières.)

Le cas d'utilisation *Conserver des documents* («*Unterlagen erhalten*») peut encore être subdivisé (voir figure 6) en d'autres cas d'utilisation : le cas d'utilisation *Migrer des documents* à proprement parler, qui assure, par la migration du format, la disponibilité des données dans un format de fichier pérenne, et les deux cas d'utilisation déjà mentionnés *Tenir à jour le catalogue des formats* et *Gérer les conversions*. Le cas d'utilisation *Tenir à jour le catalogue des formats* reconnaît et consigne la nécessité d'une éventuelle migration des données. Si c'est le cas, une migration est prévue pour le format obsolète dans *Gérer la conversion de format*. La migration à proprement parler des documents peut commencer sur cette base.

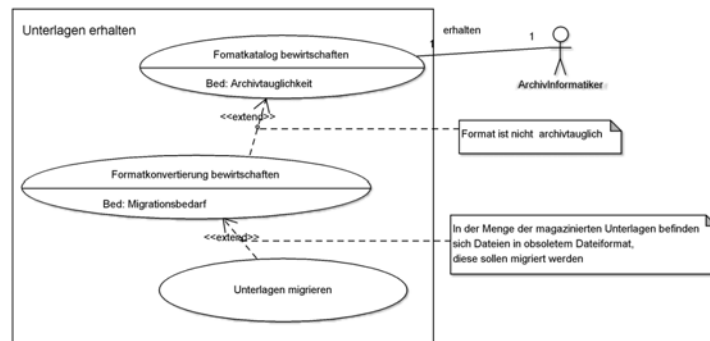


Figure 6 : Diagramme du cas d'utilisation Conserver des documents

2.4.1.1 Migrer des documents

L'activité *Migrer des documents* («*Unterlagen migrieren*») suppose une infrastructure qui est fournie par le cas d'utilisation *Gérer la reconnaissance des formats*. Cette structure de reconnaissance des formats, sous forme d'un tableau, est indépendante des Archives et peut être mise à disposition par un service central. «*Pérenne*» («*archivtauglich*») signifie que le format correspond à la politique des Archives concernées en matière de formats de fichiers ou à une politique élaborée au niveau de la Confédération. Un format pérenne est considéré comme *obsolète* quand il n'est plus recommandé de convertir des fichiers dans ce format. Les fichiers déjà présents sous ce format ne sont cependant pas convertis tant que ce format est encore considéré comme «*pérenne*».

L'ensemble des fichiers à migrer est l'ensemble des fichiers qui sont présents dans l'espace de stockage numérique et dont le format n'est plus pérenne. Les paquets d'information concernés peuvent être identifiés à partir du journal des entrées, qui indique, entre autres informations, le format de chaque fichier figurant sur l'espace de stockage.

La pérennité des fichiers présents dans un paquet d'information conservé sur l'espace de stockage numérique est éprouvée au moment de la migration. Il y a besoin, lors de la conversion à proprement parler, du format source et du format cible, ainsi que d'un outil de conversion. Le cas d'utilisation *Gérer la conversion des formats* met les éléments nécessaires à disposition.

L'historique de la migration est consigné dans les métadonnées techniques. Le paquet d'information actualisé est réinscrit dans l'espace de stockage numérique. Les actions effectuées sur le paquet d'information sont consignées dans le journal des mouvements.