

Le coût d'une plate-forme d'archivage électronique

Résultats d'une étude lancée par la direction des
Archives de France

F. Banat-Berger
Présentation Groupe PIN
20 octobre 2005

Plate-forme d'archivage électronique

- Dans un premier temps, ont été définies des sources d'informations candidates à l'archivage électronique :
 - Données (extraction à partir de bases de données)
 - Documents issus d'applications de Gestion électronique de documents (GED) : exemple des dossiers de titres de séjours dans les préfectures
 - Flux de données et documents issus des téléprocédures en cours de constitution (avec le développement de l'administration électronique)

Plate-forme d'archivage électronique

- Des hypothèses ont été établies concernant les volumes à archiver par an, concernant ces sources d'informations ; le type de transmissions (annuelle, mensuelle, non en temps réel)
- Des scénarios ont été définis visant à cerner plusieurs types d'organisations :
 - Un scénario dédié à un seul service producteur qui assure lui-même son archivage ou dispose d'un service d'archive interne (exemple d'un conseil général)
 - Un scénario pour une plate-forme dédiée à un ensemble de services producteurs locaux (type un service d'archives départementales)
 - Un scénario pour une plate-forme nationale dédiée à l'ensemble des services centraux et déconcentrés relevant d'un même ministère
 - Un scénario pour une plate-forme nationale dédiée à un ensemble de collectivités de même type (exemple des conseils généraux)
 - Enfin une plate-forme nationale pour l'ensemble des ministères (type Archives nationales)

Les volumes

- A partir de là, des volumes sur 10 ans (en incluant le fait qu'il faut doubler le volume, en raison des copies effectuées pour chaque objet archivé) ont été extrapolés, en calculant les volumes à archiver par an, suivant la source, le nombre moyen de transmissions (par exemple, pour une GED, une par an et pour une téléprocédure, une par mois) et le nombre de services versants (exemples : un pour le scénario 1 et 20 pour le 2)

Les volumes

- On a rajouté à ce nombre, dans les scénarios concernés, le volume représenté par la numérisation image des fonds papier, qui se pratique de plus en plus dans les services d'archives
- On arrive ainsi à des volumes qui restent modestes : 1 TO pour le scénario 1, 10 pour le 2 et le 4, 20 pour le 3 et 100 pour le 5

Les fonctionnalités

- Elles ont été étudiées sur la base du modèle OAIS
 - Exemple du fonds numérisés des dossiers de titres de séjour en préfecture
 - -Objet-contenu : fichiers JPG correspondant aux images numérisées relatives à un dossier
 - -Information de représentation : documentation associée au format JPG...
 - -Information de pérennisation : informations générales relatives aux dossiers archivés (numéros des dossiers, noms et prénoms des personnes concernées, nationalité, date de création du dossier, ...)
 - -Information d'empaquetage : référence des images associées à un même dossier
 - -Information descriptive : données mises à la disposition des utilisateurs pour identifier un dossier (numéro de dossier, nom de l'individu, nationalité ...)

Les fonctionnalités

- **Délibérations transmises par les collectivités aux préfetures (contrôle de légalité FAST)**
 - -Objet-contenu : fichiers PDF (ou base64) correspondant aux délibérations transmises et informations de transmission et de signature associées
 - -Information de représentation : documentation associée aux formats PDF et Base 64 ...
 - -Information de pérennisation : informations générales relatives aux délibérations archivées (collectivité émettrice, nom de l'utilisateur ayant réalisé la transmission, date de l'envoi, ...), empreinte de l'objet, référence de la délibération transmise
 - -Information d'empaquetage : lien vers la délibération
 - -Information descriptive : données utilisées pour identifier une délibération (date de la délibération, collectivité émettrice)

SIP et AIP

- Il est proposé qu'outre l'élément Archive (objet+empreinte+description) encapsulé, qui est conservé, soient extraits de l'enveloppe XML, également pour conservation, les fichiers constituant l'objet qui seront remis, le cas échéant, dans leur format binaire (PDF, PNG) et qu'ils soient ainsi conservés.

Fonctionnalités

- A partir de là, les fonctionnalités de la plate-forme ont été spécifiées, par grands processus (versement, stockage, gestion des données descriptives, restitution) et fonctions transverses (administration technique et fonctionnelle, pilotage, veille technologique et juridique, projets d'évolutions et de migrations)
- Les tâches décomposées avec la détermination d'un temps moyen par tâche et un coût horaire pour des agents de catégorie A, B et C
- Des estimations différenciées étant réalisées suivant qu'on disposait de plates-formes manuelles ou automatisées ; suivant que les transmissions se font sur supports amovibles ou par réseau

Fonctionnalités

- Fonction versement :
 - elle inclut les mécanismes de préparation (formatage des archives suivant le schéma XML), transmission, contrôle, rejet, complément d'informations et traitements pour une intégration dans le dispositif de stockage d'une part, et celui de gestion des données descriptives d'autre part.
- Fonction stockage :
 - Elle gère l'ensemble des services liés à la conservation des archives jusqu'à éventuellement leur destruction (y compris les opérations de contrôle, régénérations, migrations de supports ou de formats, restaurations..)

Fonctionnalités

- La fonction « gestion des données descriptives » :
 - Elle gère la conservation, la mise à disposition et la mise à jour des informations descriptives sur les données et documents conservés par la fonction « stockage ». Elles sont issues de l'élément description qui accompagne l'objet Archives lors du versement

Fonctionnalités

- La fonction “consultation et restitution” :
 - La restitution des documents peut être à destination du service qui les aura confiés au service d'archives (soit directement, soit par l'intermédiaire d'une plate-forme d'échanges) ou à destination d'un citoyen intéressé.
 - Elle doit garantir l'accès en ligne aux données descriptives
 - La restitution du contenu des archives s'effectuera généralement en mode asynchrone, soit par réseau (comme pour la réception des flux XML), par messagerie ou par envoi de supports amovibles.

Fonctionnalités

- Un contrat de service à établir entre le service d'archives et les services versants
 - conditions générales
 - conditions particulières (établies lors de l'initialisation d'un nouveau processus d'archivage)

Processus

- Processus A : versement d'un lot d'archives
- Processus B : stockage et conservation des informations archivées
- Processus C : consultation et restitution des archives
- Processus D : initialisation d'un nouveau processus d'archivage

Processus A

- Système de gestion des versements. Se compose de :
 - application de gestion des versements
 - base de données recensant l'ensemble des transmissions effectuées par les services versants
 - interface de type Web permettant à ces derniers de référencer les transmissions qu'ils s'appêtent à effectuer
 - d'un service de transmission par réseau (transmission des données, chiffrement et intégrité des données). Plusieurs cas de figure suivant que les transmissions se font par réseau ou non, et dans ce cas sont manuelles ou automatiques

Processus A

- d'un annuaire utilisateurs (authentification)
- d'un ou plusieurs postes de travail pour la prise en charge et le contrôle des transmissions et versements
- d'un espace de stockage temporaire des fichiers, avant transmission vers la plateforme de stockage

Processus A

- A la fin du processus, les fichiers sont mis à disposition du système de stockage (extraction de l'objet contenu et génération éventuellement des fichiers dans leur format binaire d'origine)
- Les informations descriptives sont intégrées dans la base archives (et complément d'indexation éventuel)

Processus A

- A partir de ces processus, détermination par sous-processus, des différentes tâches associées (soit à la charge du service producteur, soit à la charge du service d'archives) et étude économique en fonction d'un temps passé à chaque tâche

Processus B

- Pré-requis
 - chaque contenu est conservé sous sa forme encapsulée et sous sa forme non encapsulée (cas particuliers des fonds numérisés)
 - tout contenu archivé est conservé sur 2 sites géographiques différents
 - en utilisant deux types de supports différents

Processus B

- On aura des plates-formes manuelles (contenants sur rayonnages) ou automatisées (juke-box, librairies...), ce qui aura évidemment un impact notamment sur les coûts humains
- la plate-forme quelle qu'elle soit, comprendra un logiciel de gestion du stockage permettant de gérer la localisation physique ou électronique de chaque contenu archivé, d'organiser et d'assurer la traçabilité des opérations tout au long du cycle de vie des fichiers
- des postes de travail avec les logiciels et périphériques nécessaires aux contrôles des supports et à leur régénération

Processus C

- Interface de consultation, de recherche et de commande des fichiers archivés exploitant les informations conservées dans la base archives
- logiciel de gestion des commandes passées par les utilisateurs

Processus C

- Espace de stockage temporaire des fichiers restitués
 - contrôle avant mise à disposition,
 - accès aux fichiers en cas de consultation à partir du réseau des archives,
 - préparation de l'inscription des fichiers sur supports amovibles si nécessaire,
 - préparation de la transmission des fichiers par réseau si transmission par réseau)

Processus C

- Postes de travail équipés des matériels et logiciels nécessaires pour contrôle et production de supports amovibles
- service de transmission sécurisée par réseau (transmission, intégrité)

Processus D

- Implémenter les services de versement et de restitution au sein d'un nouveau service versant
- définir et mettre en œuvre un nouveau flux de versements au sein de ce service, concernant une nouvelle source d'information à archiver

Processus D

- Conception/customisation de la solution fonctionnelle (concordance entre les métadonnées métier et les métadonnées EAD / définition des critères de recherche)
- Développement sur cette base
- Recette
- Déploiement et démarrage
- Formation et conduite au changement associées

Fonctions transverses

- Administration de la plate-forme
- Pilotage de la plate-forme et du service d'archives (mise à jour du PAQ, production de statistiques d'utilisation, gestion des ressources humaines, gestion de budget et des dépenses, planification des tâches)
- Gestion de la relation avec les services producteurs
- Veille technologique et juridique

Solutions techniques

- Composants applicatifs
 - annuaire utilisateurs
 - application de gestion des versements
 - base archives (base de données, fonction d'importation des données en XML/EAD, interface utilisateurs pour les services d'archives, interface utilisateur de type Web pour les producteurs et le public)

Composants applicatifs

- Application de gestion des commandes
 - interface utilisateur permettant la création d'une commande
 - base de données recensant l'ensemble des commandes reçues
 - interface utilisateur pour le service d'archives (consultation et modification du bon de commande, contrôle et préparation des lots de restitution, facturation éventuellement)

Composants applicatifs

- Base de connaissances
 - base documentaire recensant l'ensemble des informations associées aux différentes applications sources faisant l'objet de versements d'archives
 - métadonnées sur les applications sources
 - documents de référence comme le PAQ, les contrats de service, les tableaux de gestion, les informations techniques sur les formats et supports...

Composants applicatifs

- Infrastructure d'hébergement des composants applicatifs
 - plusieurs serveurs permettant de les héberger (hypothèse d'une architecture 3 tiers : base de données, applications, interface utilisateur)
 - plusieurs postes de travail

Composants stockage

- Supports de versement et de restitution
- Supports de stockage et de sauvegarde
 - disques optiques
 - bandes magnétiques
 - disques magnétiques
- Plates-formes de stockage
 - archivage sur mobilier
 - archivage sur périphérique de stockage (avec logiciel de gestion de stockage)

Composants réseau

- Réseau interne au service des archives
- Réseau de transmission services producteurs/services d'archives (au niveau des versements / des restitutions)
- Réseau de stockage

Etude économique

- Distinction entre coûts nationaux et coûts « plate-forme »
- Distinction entre coûts initiaux et coûts annuels d'exploitation (ces derniers se retrouvant et au niveau national et au niveau « plate-forme »)
- Distinction entre coûts internes et externes

Plate-forme d'archivage électronique

- Analyse économique basée sur l'hypothèse suivante :
 - Le développement (et la maintenance) au niveau national, plutôt sur la base de progiciels du marché (pour lesquels des travaux importants de paramétrage devront être effectués, étant donné l'état du marché sur ce type de produits), d'une solution générique, en passant par les étapes : développement, réalisation d'un déploiement pilote, généralisation de la solution (coûts initiaux et coûts annuels d'exploitation)

Plate-forme d'archivage électronique

- Ont été également estimés les coûts spécifiques à chaque plate-forme à savoir :
 - Les coûts initiaux (acquisition de la plate-forme d'exploitation, installation de la solution générique adaptée, coûts interne de démarrage auprès des services producteurs, démarrage des applications sources.)
 - Les coûts d'exploitation annuels, suivant les processus (prise en charge des versements, coûts de gestion du stockage, coût des restitutions et consultations, coûts de maintenance applicative et technique, coût des fonctions transverses)
 - Des synthèses ont été réalisées scénario par scénario ainsi que des comparatifs des coûts au TO

Plate-forme d'archivage électronique

- Concernant les supports et les périphériques de stockage, l'étude préconise :
 - l'utilisation de bandes magnétiques type LTO 2 ou 3
 - Celle de disques magnétiques type SATA s'il existe des exigences élevées en terme de restitution
 - Écarte l'utilisation des DVD-R (hormis pour de très petits volumes) en raison de leur cherté
 - Préconise l'utilisation de plates-formes automatisées uniquement à partir d'un certain volume (plus de 500 MO par an au moins). Ensuite, jusqu'à 2 TO de données, tout dépend des moyens humains

Plate-forme d'archivage électronique

- Pour ce qui est de la sécurité et des réseaux, l'étude préconise :
 - Un accès internet classique avec des débits de 1 et 2 Mb montants, au minimum
 - Transmissions par SSL (HTTP ou FTP/SSL) avec par défaut uniquement des certificats serveurs, en s'appuyant sur l'utilisation d'un moniteur de transfert qui sache effectuer des reprises de transferts interrompus
 - Mise en place d'une machine dédiée gérant l'accès à l'internet et équipée d'un pare-feu (ce système permettra de gérer une zone démilitarisée DMZ)
 - Une identification/authentification par user/password avec en option, l'utilisation de clés privées et de certificats

Plate-forme d'archivage électronique

- Pour ce qui est du réseau de stockage, les systèmes type NAS ne sont prescrits que pour les plus grandes plates-formes (gérant plusieurs dizaines de TO)
- Enfin, concernant le site d'hébergement de la plate-forme, sont préconisés des onduleurs, des contrôles d'accès par badges ou clés (distincts de ceux prévus pour l'accès dans les autres pièces du site si celui-ci n'est pas dédié au service d'archives), et un contrôle d'accès supplémentaire pour la salle informatique, utilisation d'une salle blanche avec une charge au sol suffisante)...

Plate-forme d'archivage électronique : conclusions

- Dès lors que les volumes archivés augmentent, les économies d'échelle sont considérables
 - les ratios entre les coûts par TO sont systématiquement supérieurs à 20 entre le scénario 1 (un seul producteur local) et le 5 (tous les ministères) que ce soit pour les coûts initiaux externes ou les coûts d'exploitation
 - elles sont de 1 pour 100 pour les coûts externes initiaux dans le cas d'une plate-forme manuelle et de 1 pour 33 dans celle d'une plate-forme automatisée:
- l'étude préconise par conséquent de privilégier les plates-formes d'une certaine ampleur

Plate-forme d'archivage électronique : conclusions

- Les coûts internes les plus importants sont ceux afférents au démarrage d'un processus d'archivage avec un service producteur (par service producteur / par nouvelle application) en catégories A pour le démarrage et en catégories B pour l'exploitation, que ce soit pour le service d'archive ou pour le producteur qui devra adapter son application pour le transfert au format défini par le schéma XML de versement vers la plate-forme d'archivage
- Par conséquent, il vaut mieux privilégier les gros volumes de versements et limiter autant que possible le nombre de producteurs différents, si on ne dispose pas des ressources humaines suffisantes
- De même, il est préconisé d'étoffer l'échelon national (DAF) afin d'apporter une aide à la modélisation des versements par grandes catégories de documents, que l'on retrouve dans tous les sites (type marchés publics)

Perspectives

- Dans les mois à venir, la DAF a l'intention en étroite collaboration avec l'ADAE :
 - De faire connaître et de diffuser l'étude (communications diverses par le biais de l'ADAE ou d'autres instances comme celle-ci
 - De contribuer à l'alimentation des référentiels sécurité et interopérabilité concernant l'archivage
 - D'écrire un cahier des charges pour lancer le pilote, embryon de la solution générique que préconise l'étude, et à travailler avec des sites tests pour la généralisation (collectivités intéressées comme par exemple les Yvelines, centre des Archives nationales de Fontainebleau et du futur centre de Pierrefitte....

Perspectives

- D'alimenter les référentiels de l'action 103 d'ADELE :
 - appel à commentaires pour le standard d'échanges déjà lancé
 - référentiels sur les processus et l'architecture technique
 - sur le cycle de vie du document (travail en commun avec l'APROGED)