

Paris, le 13 octobre 2005

La directrice des Archives de France

à

Mesdames et Monsieur les directeurs des centres des Archives nationales

Mesdames et Messieurs les présidents des conseils généraux (archives départementales)

Mesdames et Messieurs les maires (Archives communales)

Mesdames et Messieurs les présidents des conseils régionaux (Archives régionales)

Note d'information DITN/RES/2005/0007 relative à une étude relative à la réalisation de plate-formes d'archivage électronique

La direction des Archives de France a fait réaliser par la société Parker-Williborg, une étude sur le coût de plates-formes d'archivage électronique destinées soit aux administrations (administrations centrales, administrations déconcentrées, collectivités territoriales, établissements publics) soit aux services publics d'archives.

Cette étude s'inscrit dans le contexte du développement de l'administration électronique et de la dématérialisation progressive de processus entiers administratifs, qui amène à devoir assurer la conservation pérenne de ces nouvelles sources, tant pour des raisons de preuves et de droits que pour l'intérêt patrimonial qu'elles peuvent représenter.

Elle entre dans l'action 103 Adele sur l'archivage électronique et le cycle de vie du document, qui prévoit d'une part, l'élaboration de référentiels (schéma XML pour les échanges vers une plate-forme d'archivage, processus fonctionnels, architecture technique et volet archivage du référentiel de sécurité..) et d'autre part, la mise en oeuvre d'une plate-forme d'archivage pour les Archives nationales et un appui pour les collectivités territoriales.

L'étude se compose de trois livrables et d'une synthèse : un état des lieux, une étude des différents scénarios ainsi qu'une étude comparative des coûts tant initiaux que d'exploitation, externes

(maîtrise d'oeuvre et assistance à maîtrise d'ouvrage, maintenance, veille..) et internes (coûts des personnels).

Scénarios et volumétries

Ont été définies dans un premier temps, les sources candidates à l'archivage électronique : données extraites des différentes bases de données produites par les administrations, documents issus d'applications de gestion électronique des documents (dans le cadre des programmes de numérisation qui se développent dans un certain nombre d'administrations), et enfin flux de données et documents issus des téléprocédures en cours de constitution avec le développement de l'administration électronique (marchés publics, contrôle de légalité...). Des hypothèses ont été formulées concernant les volumes à archiver par an et par type de sources, le nombre de transmissions (annuelle ou mensuelle).

Ces hypothèses ont permis d'élaborer un certain nombre de scénarios couvrant différents types d'organisations :

- une plate-forme dédiée à un seul service producteur qui assure lui-même son archivage ou dispose d'un service d'archive interne (exemple d'une grande municipalité, d'un conseil général) : scénario 1
- une plate-forme dédiée à un ensemble de services producteurs locaux (comme les services d'archives départementales ou certains gros services d'archives municipales) : scénario 2
- une plate-forme nationale dédiée à l'ensemble des services centraux et déconcentrés relevant d'un même ministère : scénario 3
- une plate-forme nationale dédiée à un ensemble de collectivités de même type (conseils généraux, conseils régionaux) : scénario 4
- enfin une plate-forme nationale pour l'ensemble des administrations centrales de l'Etat (Archives nationales) : scénario 5.

Des volumétries sur 10 ans ont alors été extrapolées, en doublant ces hypothèses dans la mesure où l'ensemble des ressources doivent être conservées, pour des raisons évidentes de sécurité, en deux exemplaires. Ont été inclus dans ces chiffres, pour les Archives départementales et municipales, le volume induit par la numérisation en cours des fonds papier (voir le catalogue des fonds numérisés). On atteint ainsi, suivant les scénarios, des estimations allant de 1 TO (pour le premier scénario à 10 (scénarii 2 et 4), 20 pour le scénario 4 et 100 pour le dernier.

Pré-requis

Une préoccupation forte et légitime concernant l'intégrité et plus généralement la fiabilité et la fidélité des documents, est apparue avec le développement de l'électronique. La réflexion des professionnels de l'information à ce sujet, notamment des archivistes et documentalistes (tant au niveau des institutions, que des associations professionnelles ou encore des instituts de recherche) a conduit à préférer les réponses organisationnelles et fonctionnelles pour tenter de traiter cette question complexe, aux réponses purement techniques (signatures électroniques reposant sur des technologies cryptographiques, emploi de supports dits WORM).

Ainsi, dans le domaine des archives publiques, la fiabilité du processus d'archivage repose sur les missions traditionnelles des établissements en charge de l'archivage, qui sont chargés de restituer en se basant sur les règles de la profession et leurs savoirs-faire notamment en matière de description et de conservation, les documents qui leur ont été confiés. Ceci signifie une intégration du cycle de vie du document, l'enregistrement de tout événement intervenu concernant ce document, la prise en compte de son contexte de création, de production puis de versement pour archivage.

Dans le domaine électronique, le paradoxe qui consiste à assurer la préservation de données numériques par définition dépendante d'un environnement matériel et logiciel de plus en plus complexe et évolutif sur le très court terme, implique de travailler sur la base des normes et standards existants dans le domaine des formats de représentation des données et dans le domaine de la description ; de documenter toute action en privilégiant la préservation à long terme d'un contenu sémantique sans tenter de préserver l'intégrité physique des fichiers numériques initiaux (les chaînes de 0 et de 1), impossibles à stabiliser sur le moyen et long terme.

En effet, les seules possibilités techniques actuellement offertes sont les migrations de supports et de formats, qui affecteront tôt ou tard, l'intégrité physique d'un train de bits. Dans ce cas, il faut se reposer sur les processus métier de l'archivage qui, seuls, peuvent garantir la qualité des opérations réalisées et celle des restitutions effectuées.

Il en ressort que :

L'institution en charge des archives publiques n'aura pas à vérifier les signatures originelles des documents qui lui sont confiés (cette vérification devant s'effectuer en amont, par les producteurs de l'information).

Le contrôle d'intégrité se fait au moment de la réception (contrôle des empreintes générées par le producteur, sur le lot à archiver. Elle peut également se faire au moment de la réception puis à chaque communication, sur le contenu de l'information à archiver, une empreinte ayant été également générée par le producteur à ce niveau-là, au moment du versement

Si une migration de format doit s'effectuer, un nouveau contenu archivé sera produit et une nouvelle empreinte générée.

L'institution en charge des archives publiques ne prend en charge que des documents décryptés, une valeur ajoutée de l'archiviste étant apportée notamment pour la description des archives versées, dont il importe par conséquent de connaître et maîtriser le contenu et le contexte de production

Les supports WORM ne seront pas à priori privilégiés, la preuve ne reposant pas sur une technologie particulière liée à un support mais à un processus dans son ensemble

Fonctionnalités de la plate-forme d'archivage électronique

Dans un second temps, ont été spécifiées les fonctionnalités de la plate-forme, par grands processus (préparation et prise en charge des versements, stockage, gestion des données descriptives, restitution) et par fonctions transverses (administration de la plate-forme, pilotage, veille technologique et juridique, projets d'évolutions et de migrations). Les tâches ont été ensuite décomposées avec la détermination d'un temps moyen par tâche et un coût horaire des agents de catégories A, B ou C. Des estimations différentes ont été dressées suivant le choix d'une plus ou moins grande automatisation des processus, l'utilisation pour les transmissions, de réseaux ou de supports amovibles, la mise en place de plates-formes manuelles (supports sur rayonnages) ou automatisées (juke-box, librairies, baies de disques...).

La fonction « versement » inclut tous les mécanismes de préparation (les fichiers à archiver étant encapsulés dans une enveloppe XML, suivant un format déterminé), transmission, contrôle, rejet, complément d'information ainsi que tous les traitements de ces informations pour une intégration dans le dispositif de stockage des contenus et celui de gestion des données descriptives.

La fonction « stockage » gère l'ensemble des services liés à la conservation des archives à partir du moment où celles-ci sont mises à disposition par la fonction « versement » jusqu'à leur destruction s'il y a lieu (y compris les opérations de contrôle, régénérations, migrations de supports ou de formats, restaurations...).

La fonction « gestion des données descriptives » gère la conservation, la mise à disposition et la mise à jour des informations descriptives disponibles sur les données et documents conservées par la fonction « Stockage ». Elles sont issues des métadonnées de description qui accompagnent les contenus d'information à archiver, lors du versement.

Enfin, la fonction « consultation et restitution ». Cette restitution peut se faire à destination du service qui aura confié ses archives à la plate-forme d'archivage, ou à destination du « grand public » pour les documents librement communicables. Le système doit garantir l'accès en ligne aux données descriptives et un accès généralement a-synchrone aux contenus eux-mêmes, soit par réseau, par messagerie ou par envoi de supports amovibles.

Préconisations techniques

Ces fonctionnalités conduisent à définir les différents composants applicatifs de la plate-forme : annuaire utilisateur, application de gestion des versements, base archives (base de données gérant les informations descriptives, fonction d'importation des données descriptives reçues lors du versement, interface utilisateurs pour les gestionnaires de la plate-forme, interface utilisateurs de type Web), application de gestion des commandes, base de connaissances (base documentaire). Ceci nécessite la mise en place de plusieurs serveurs et de postes informatiques adéquats.

Concernant les supports de stockage et les périphériques, l'étude recommande l'utilisation de bandes magnétiques type LTO2 ou LTO3 (et proscrit, excepté pour de petits volumes, celle de supports optiques dont le coût par rapport aux volumes à stocker, est prohibitif). Est recommandée, l'utilisation de disques magnétiques type SATA en cas d'exigences élevées en matière de restitution.

Les plates-formes automatisées ne sont recommandées qu'à partir d'un certain volume (au moins 500 MO par an). De même, pour le réseau de stockage, les systèmes type « Network Attached Storage » (NAS) ne sont prescrits que pour les plus grandes plates-formes. La solution plus onéreuse que constitue le réseau « Storage Area Network » (SAN) ne semble pas pertinente, dans la mesure où les niveaux de disponibilité et de temps d'accès ne sont pas critiques.

Enfin, concernant le site d'hébergement de la plate-forme, sont préconisés l'utilisation d'onduleurs, celle de contrôles d'accès par badges ou clés qui soient distincts de ceux pour l'accès dans les autres pièces du site si celui-ci n'est pas dédié au service d'archives, un contrôle d'accès supplémentaire pour la salle informatique, la mise en place d'une salle dite « blanche » avec une charge au sol suffisante...

Pour ce qui est de la sécurité et des réseaux, sont recommandés un accès internet classique avec des débits de 1 et 2 Mb montants, au minimum ; la transmission en utilisant une couche de sécurité spécifique (SSL) qui permet de s'assurer de l'intégrité de la transmission, en s'appuyant sur l'utilisation d'un moniteur de transfert qui sache effectuer des reprises de transferts interrompus ; la mise en place d'une machine dédiée gérant l'accès à Internet et équipée d'un pare-feu, ce système permettant ainsi de gérer une zone démilitarisée type DMZ ; enfin, une identification/authentification classique basée sur l'utilisation du couple « user/password » avec en simple option, l'utilisation de clés privés et de certificats.

Etude économique et dispositifs organisationnels

Actuellement, sur la marché, n'existe aucun produit répondant à l'ensemble des fonctionnalités requises notamment pour la partie prise en charge des versements. En revanche, commencent à

émerger quelques solutions logicielles répondant à un certain nombre des fonctionnalités demandées.

Par conséquent, l'analyse économique a été basée sur l'hypothèse suivante, à savoir le développement au niveau national, sur la base de logiciels du marché (pour lesquels un travail d'intégration important sera demandé), d'une solution générique en passant par les étapes suivantes : développement, réalisation d'un déploiement pilote, généralisation de la solution (coûts initiaux et coûts annuels d'exploitation).

Il semblerait plus avantageux de choisir le système de licences plutôt qu'un développement spécifique d'un bout à l'autre notamment pour des raisons économiques : c'est seulement dans le cas de déploiement de multiples plates-formes (plus de 10), que le développement spécifique pourrait se justifier.

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

Il en ressort que dès lors que les volumes archivés augmentent, les économies d'échelles sont considérables : les ratios entre les coûts par TO sont systématiquement supérieurs à 20 entre le scénario 1 (un seul producteur local) et 5 (Archives nationales) que ce soit pour les coûts initiaux externes ou les coûts d'exploitation. Il est par conséquent préconisé d'encourager les plates-formes d'une certaine ampleur.

Les coûts internes les plus importants sont ceux afférents au démarrage d'un processus d'archivage avec un service producteur (nouveau producteur/nouvelle application) avec une forte charge sur les catégories A au démarrage et sur les catégories B pour l'exploitation. Ceci vaut tant pour le service d'archives que 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 versements de gros volumes et limiter autant que possible le nombre de services producteurs différents, si on ne dispose pas de ressources humaines suffisantes.

De même, l'étude préconise fortement, dans un souci de mutualisation, de renforcer l'échelon central (Direction des Archives de France) afin d'apporter un aide à la modélisation des versements par grandes catégories de documents, que l'on retrouve sur tous les sites (exemple des marchés publics).

Perspectives

En relation étroite avec l'ADAE, la direction des Archives de France envisage dans les prochains mois :

- de communiquer sur sa politique et sur l'étude dans le cadre des rencontres ADELE de l'ADAE et d'ateliers thématiques

- de publier plusieurs référentiels en matière d'archivage : schéma XML pour le format des données versées (un appel à commentaires vient d'être publié sur le site de l'ADAE) ; processus d'archivage (fonctionnalités) et architecture technique (sur la base des résultats de l'étude) ; cycle de vie du document. Pour chaque thème, des appels à commentaires et des ateliers thématiques seront organisés

de participer à la rédaction des référentiels interministériels sur l'interopérabilité et la sécurité (concernant ce dernier domaine, une étude vient d'être lancée par le DCSSI, à laquelle la DAF et l'ADAE participent)

de lancer en 2006 une étude sur les différents dispositifs organisationnels possibles en matière d'archivage électronique notamment pour les collectivités territoriales : état de lieux sur la base du Droit existant, évolutions possibles.

de lancer un appel d'offres sur la base d'un cahier des charges qui sera rédigé au début de l'année 2006, pour la réalisation d'un déploiement pilote pour la plate-forme d'archivage électronique (sur le site du centre des archives contemporaines à Fontainebleau) avant déploiement sur le futur site des Archives nationales à Pierrefitte.

La directrice des Archives de France

Martine de BOISDEFFRE