

**Compte rendu de colloques : de Berne (archivage des bases de données) à Louvain (les archives électroniques, un défi pour l'avenir ?)**

L'archivage des bases de données : Erpanet Workshop de Berne, 9-11 avril 2003

L'une des missions d'Erpanet, réseau consacré à l'archivage électronique financé par la Commission européenne, est de contribuer à la formation continue des archivistes par la tenue de séminaires.

Après le séminaire de janvier 2003 consacré aux politiques d'archivage et tenu à Fontainebleau, l'atelier d'avril 2003 portait sur un type de document électronique très diffusé et dont la préservation n'est pas sans poser d'épineux problèmes. L'Europe du nord et les Etats-Unis étaient particulièrement bien représentés parmi les intervenants et parmi les participants ; la plupart étaient issus de services publics d'archives. L'atelier était structuré en quatre parties où les présentations de cas étaient très bien représentées.

Le séminaire, de deux jours et demie, était structuré en quatre séquences d'inégales longueurs : après un exposé introductif sur les enjeux de l'archivage des bases de données, un grand nombre de présentations ont été consacrées aux expériences pratiques d'archivage des bases de données, puis deux demies journées ont été consacrées, pour l'une aux méthodes proprement archivistiques d'évaluation et de description, pour l'autre, aux problèmes posés par la réutilisation des données.

Il n'est pas question de donner ici un résumé détaillé du séminaire (voir les présentations utilisées à cette occasion sur le site [www.erpanet.org](http://www.erpanet.org) sous la rubrique " products "), dont bon nombre d'exposés avaient un caractère technique assez prononcé. Il semble plus intéressant de dégager les problèmes posés par les différentes opérations archivistiques lorsqu'elles portent sur un tel type documentaire, et de mettre en regard les différentes approches proposées, c'est pourquoi nous commencerons par les stratégies de collecte, avant de parler de la description des données et de terminer par la mise en valeur.

La collecte : quelques approches et leurs implications

Trois types de stratégies d'archivage semblent exister, pour l'instant, qui ne sont pas forcément exclusives les unes des autres.

La conservation " à plat " sous la forme de fichiers textuels (à l'origine, au format ASCII) est la stratégie la plus ancienne, qui consiste à isoler les données de leur environnement logiciel. Appliquée aux Archives nationales, elle est aussi en usage dans d'autres institutions, notamment aux Etats-Unis ([www.archives.gov](http://www.archives.gov)) et en Grande-Bretagne ([ndad.ulcc.ac.uk](http://ndad.ulcc.ac.uk)). L'inconvénient de cette approche est qu'elle ne permet pas une consultation facile des données, qui ne sont interprétables qu'une fois la base de données reconstituée à partir des métadonnées.

La conservation sous forme de fichiers XML des données à caractère historique. Le choix du XML comme format de conservation doit faciliter la réutilisation, tout en évitant les problèmes liés aux changements de structure des tables des bases de données. Le projet DAVID, développé en Belgique en région flamande, s'appuyait très largement sur un choix de ce type. Aux archives municipales d'Anvers, la collecte des données issues des applications de gestion passe par leur conversion en XML : soit par transformation des bases de données relationnelles en bases de données XML (un seul fichier XML), soit par conservation des tables des bases de données sous la

## Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 13 novembre 2003

forme de fichiers XML (autant de fichiers XML que de tables dans la base de données relationnelle d'origine). Aux Pays-Bas comme en Belgique, le choix a été fait, dans le cadre du projet Digital Preservation Testbed (voir ci-dessous l'article consacré à ce projet), de mettre au point des outils de conversion en XML qui fonctionnent à partir des logiciels de base de données les plus courants. Cette manière de traiter le problème suppose de faire évoluer les outils de conversion en XML au rythme des évolutions des bases de données sources. Idéalement, il faudrait un module d'export en XML depuis l'application d'origine. En pratique, il semble que cette fonction soit encore peu répandue.

La conservation de la base de données en tant que témoins de toutes les activités d'un milieu social implique la conservation du plus de fonctionnalités et de données possibles de la base de données initiale : une fois qu'une application a été évaluée comme intéressante du point de vue archivistique, il n'est plus question d'opérer des tris dans les données qu'elle traite, sauf de façon marginale. Tel est le choix effectué par les archives fédérales suisses ([www.admin.ch/bar](http://www.admin.ch/bar)), qui ont opté pour un ensemble d'applications, appelé SIARD (Software Invariant Archiving of Relational Databases ; voir article sur ce projet ci-dessous), qui permettent à la fois l'analyse, l'intégration et la restitution d'une très grande partie des fonctions et des données produites par les bases de données d'origine. Ce choix apparaît probablement comme le plus coûteux, sans doute aussi comme le plus exigeant du point de vue technique.

Indépendamment de la stratégie retenue, les bases de données posent un problème particulier à l'archiviste qui entend les traiter, celui de leur mise à jour. En effet, les bases de données peuvent traiter les modifications de leur contenu de deux manières différentes :

soit par écrasement du contenu antérieur, lorsque la connaissance de celui-ci importe peu au service utilisateur (par exemple, les logiciels de gestion d'archives ne gèrent pas les adresses successives des lecteurs inscrits) ;

soit par constitution d'un historique des modifications apportées à une affaire (modifications apportées au registre du commerce et des sociétés tenu électroniquement).

Selon les cas, il faudra donc choisir entre la conservation d'une " photographie " de la base de données à un moment précis ou l'archivage d'affaires dont le délai d'utilité administratif est clos.

Les métadonnées : de la collecte à l'utilisation

Quel que soit le choix adopté, l'archivage s'accompagne de la collecte ou de la reconstitution d'une documentation sur les données. La collecte des métadonnées, sur support papier ou électronique, est presque unanimement reconnue difficile, sauf apparemment dans certains pays nordiques, à la fois riches, centralisés et peu peuplés. La difficulté de la collecte tient le plus souvent à une gestion peu attentive de leurs documents par un certain nombre de services informatiques.

En Norvège, les Archives nationales ont mis au point une DTD (ADDMMML) pour les métadonnées, qui est antérieure aux normes internationales en la matière, mais qui est en cours d'évolution pour leur être compatible. C'est en utilisant cette DTD que toutes les données collectées sous forme électronique sont décrites.

Du point de vue de l'archiviste, l'expérience menée dans le cadre du NDAD semble cependant plus intéressante, puisqu'il s'agit de l'application de la norme ISAD (G) à la description des documents

## **Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 13 novembre 2003**

électroniques : une telle option, même si elle suppose d'aménager un peu la norme (comme il apparaît à la lecture des instruments de recherche publiés sur le site indiqué plus haut), a l'avantage de la continuité avec les pratiques existant pour le papier. En l'absence de métadonnées cohérentes, la pratique adoptée par le NDAD est de conserver malgré tout les données, tout en signalant au chercheur leur incohérence et la difficulté d'expliquer l'incohérence.

C'est au sein des Archives nationales américaines que l'utilisation des métadonnées pour vérifier l'intégrité des fichiers transmis est la plus poussée. Le système AERIC (Archival Electronic Records Inspection and Control System) permet la vérification presque totalement automatisée des quelque 2000 fichiers transmis par les administrations fédérales américaines chaque année. La quantité de données transmises justifie l'automatisation du contrôle

### **Communication et mise en valeur des données**

Les bases de données archivées sont encore relativement peu communiquées, en raison de la nature des données traitées (cas d'une très grande partie des données conservées à Fontainebleau) ou de leur caractère très récent.

L'accès en ligne est très rarement prévu, sauf dans par le NDAD, qui offre la consultation en ligne de certaines tables de quelques-unes des applications archivées. Mais même dans le cadre de cette expérience, la communication est extrêmement limitée, pour des raisons de protection de la vie privée.

Du point de vue de la réutilisation des données, c'est sans doute pour les bases de données à caractère scientifique que les mises en œuvre ont été les plus fécondes. En effet, la réutilisation permet d'éviter de reproduire des expériences parfois très coûteuses. Elle permet aussi d'utiles confrontations avec les données issues de travaux plus récents. C'est pourquoi la conservation a été pensée, dans ce contexte, en lien très étroit avec la réutilisation.

Il serait trop facile, pour conclure, de se contenter de saluer la qualité, d'ailleurs remarquable, des exposés. Voici au moins deux points qui m'ont semblé mériter réflexion.

Le titre du séminaire était " long-term preservation ", que l'on peut traduire par conservation à long terme. Or la conservation proprement dite n'a fait l'objet d'aucun exposé en particulier. Je n'y vois pas une négligence des organisateurs, mais plutôt le symptôme de l'intégration de cet aspect à toutes les étapes du traitement d'un document électronique par les archives, et particulièrement de la collecte.

L'objet qu'il faut archiver a beau être nouveau, les archivistes restent très liés à leur tradition archivistique et administrative. Les concepts ont beau être proches, bien des termes employés se révèlent trompeurs. Des comparaisons comme celles que permet ce séminaire sont d'autant plus nécessaires pour pouvoir envisager un jour des coopérations.

Les archives électroniques : un défi pour l'avenir ? Louvain-la-Neuve, 8 et 9 mai 2003

3e journée organisée par le service des Archives de l'université catholique de Louvain,. Publication des actes prévue en décembre 2003. Un résumé en est consultable sur le site : [www.ucl.ac.be/arcv](http://www.ucl.ac.be/arcv)

La préoccupation centrale de la journée d'études organisée par le service des Archives de

## **Bulletin des Archives de France sur la conservation à long terme des documents électroniques** **N° 13 novembre 2003**

l'université catholique de Louvain était l'évolution de la profession, confrontée aux évolutions technologiques et à l'émergence – désormais assez ancienne – de professions liées aux technologies de l'information, et qui peuvent parfois sembler concurrentes.

Après une conférence inaugurale de F. Burgy, dont le propos était plus optimiste que le titre ne le laisserait croire (" Les documents électroniques vont-ils avoir raison des archivistes ? "), la dizaine de communications illustre les deux faces de la présence des documents électroniques dans les services d'archives : la collecte et la production ; la communication et la mise en valeur.

Tous les intervenants ont souligné la nécessité d'un renouvellement, sinon des fondements de l'archivistique – dont la plupart des concepts de base, comme le rappelait J. M. Yante, restent valides –, en tout cas des pratiques.

De fait, l'archiviste n'est plus et ne peut plus être seul à opérer, que son travail s'effectue en amont ou a posteriori. Sans doute les compétences informatiques, parfois limitées, sont en cause ; plus profondément, étant données les implications de l'informatisation, la mise en œuvre d'un projet informatique multiplie les interlocuteurs.

De ce point de vue, l'expérience du projet DISSCO, mené au sein de la Vrije Universiteit de Bruxelles paraît exemplaire (F. Scheelings), puisque ce projet de gestion électronique des procédures et des documents relatifs à l'administration de l'établissement a associé des spécialistes de l'ingénierie des systèmes d'information, des archivistes et des sociologues.

A la Ville d'Anvers (F. Boudrez), les archivistes ont pu s'appuyer sur le projet de recherche DAVID, financé par la communauté flamande de Belgique, pour faire évoluer les instruments de travail et les pratiques des agents de la commune. L'exemple proposé ici était celui des messageries électroniques. S'il n'est pas possible de forcer les agents à verser leurs messages, il est pertinent de les y inciter et de leur procurer un outil qui permette un versement aisé (ce qui passe par un paramétrage de l'outil de messagerie électronique). Une fois le versement réalisé, il n'y a plus qu'à opérer une conversion dans un format pérenne. Un outil de conversion en XML a été développé à cette fin.

L'archivage électronique n'est pas sans poser des problèmes juridiques nouveaux du point de vue du droit. Comme le rappelait L. Verachten, la notion d'authenticité évolue, ce qui n'est pas sans susciter de nombreuses questions (d'autant plus épineuses dans un pays multilingue).

Des conflits d'intérêt se développent. H. Dekeyser étudie l'exemple du Moniteur belge, dans lequel sont publiées toutes sortes d'informations légales, nominatives ou non. Le fait de les mettre en ligne ne constitue certes pas une publication nouvelle par rapport à la version " papier ", pourtant la vie privée des personnes est autrement mise en cause. D'un autre point de vue, le principe du respect des fonds, qui permet de garantir l'intégrité des informations historiques, peut s'opposer au droit d'accès et de rectification des personnes concernées. De là la nécessité de mettre en œuvre des compromis acceptables.

Le lien entre collecte et mise en valeur n'est pas neuf. C'est même l'un des fondements de l'archivistique et la justification de services d'archives constitués. L'avènement de documents électroniques oblige cependant à établir un lien plus étroit entre toutes les étapes du cycle de vie d'un document, pour ne pas conserver " une montagne de débris électroniques ", selon l'expression du fr. F. Poswick. Plus que jamais, la conservation constitue un projet, et non

## **Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 13 novembre 2003**

seulement un stockage passif.

C'est ce que la mise en œuvre du projet IMAGO à l'Archivio di Stato de Florence démontre amplement (F. Klein). La numérisation d'actes anciens dite "Diplomatico" permet de concilier la conservation d'un ensemble prestigieux et la valorisation de documents dont la publication en ligne favorise l'exploitation selon des modalités inédites.

Si, comme il a déjà été dit, bien des concepts de l'archivistique demeurent valides pour décrire des réalités nouvelles, il n'en demeure pas moins vrai que de nouveaux termes se diffusent, reflétant les préoccupations de milieux professionnels étrangers aux archives (notamment des médecins) en matière de gestion des documents et des savoirs. L'apparition et la diffusion d'expressions "nouvelles", telles que "records management", "knowledge management", "computerized records" peut être suivie grâce à une enquête bibliométrique (en cours), menée sous la direction de M. Dorban à partir des bases de données Lisa, Inspec, Medline, Francis et Current contents.

Au total, cette journée constituait une occasion de s'interroger sur la profondeur des évolutions que connaissent nos pratiques, et surtout une belle moisson d'expériences concrètes.

Louis Faivre d'Arcier

### **Un projet suisse d'archivage des bases de données : SIARD**

Les Archives fédérales suisses ont développé un système original d'archivage des bases de données relationnelles. C'est le logiciel SIARD (Software Invariant Archiving of Relational Databases). L'originalité du système réside dans le souci d'automatiser le plus possible la collecte et de préserver les liens dynamiques des bases archivées. Dans ce but, elles ont choisi d'utiliser le format SQL-3 (mais de façon strictement conforme à la norme ISO/IEC 9075) comme format de conservation de la structure de la base de données, au côté du format "à plat" encapsulé en XML pour les données. L'application, qui a été développée en collaboration avec la société Triviadis, est capable d'analyser automatiquement les données à archiver et, après intervention d'un opérateur humain pour compléter le résultat de cette analyse, de les exporter vers le système d'archivage accompagnées des métadonnées appropriées (métadonnées elles-mêmes organisées sous XML). SIARD peut aussi recharger les données archivées dans une nouvelle base, qui permet d'effectuer des recherches. En moment, SIARD se trouve en phase de test.

Pour obtenir plus d'informations: [stephan.jaermann@bar.admin.ch](mailto:stephan.jaermann@bar.admin.ch)

### **Coopération européenne : l'agenda de Florence**

La présidence italienne de l'Union européenne a organisé à Florence les 16 et 17 octobre un colloque intitulé "l'avenir des ressources numériques du patrimoine". Outre des présentations fort intéressantes (voir ci-dessous l'expérience néerlandaise de comparaison des méthodes de conservation) ce colloque a été l'occasion de lancer "l'agenda de Florence", qui est un plan d'action de l'union européenne pour l'archivage électronique. Ce document, qui s'inscrit dans la suite logique de l'adoption, par le conseil européen d'une résolution sur la conservation du patrimoine numérique,

## **Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 13 novembre 2003**

devrait être adopté dans sa forme définitive lors de la rencontre de Parme le 19 novembre.

### **L'évaluation comparée des méthodes de conservation : l'expérience néerlandaise**

Les archives nationales des Pays-Bas ont mis au point un banc d'essai des méthodes de conservation des documents numériques. Cette expérience originale est menée dans le cadre du projet global de développement de l'administration électronique du gouvernement néerlandais.

Les tests portent sur les catégories de documents électroniques suivantes :

les documents de traitement de texte ;

les feuilles de calcul (type Excel) ;

les messages électroniques ;

les bases de données ;

Les méthodes testées sont :

la migration (entendue au sens de "conversion d'un environnement logiciel et/ou matériel vers un autre", ce qui inclut différents types de migration) ;

la conversion en XML ;

l'émulation (là aussi définie au sens large et incluant plusieurs sous-catégories) ;

Actuellement les résultats disponibles concernent la conservation des messages électroniques. Ils concluent à une plus grande efficacité de la méthode basée sur la conversion en XML. Les rapports concernant les autres documents devraient paraître d'ici la fin de l'année 2003. Enfin, et ce n'est pas le moindre intérêt de ce projet, une partie consacrée à la comparaison des coûts de conservation selon les méthodes est aussi prévue.

Pour en savoir plus :

- le site du banc d'essai néerlandais : <http://www.digitaleduurzaamheid.nl/home.cfm>

### **Une "bouteille à terre" électronique**

Il est de tradition, lors de la construction d'un bâtiment, et particulièrement lorsqu'il s'agit d'un dépôt d'archives, d'enterrer dans ses fondations un objet ou un document destiné aux générations futures.

## **Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 13 novembre 2003**

Généralement, il s'agit de pièces de monnaie ou de médailles. Nos collègues des Archives départementales d'Eure-et-Loir ont su respecter la tradition tout en faisant preuve d'esprit d'innovation : à l'occasion de la pose de la première pierre de leur futur nouveau bâtiment, ils ont en effet choisi d'enterrer (incluse dans une pierre en verre) la bande de sauvegarde des fichiers informatiques de gestion des Archives départementales.

L'histoire ne dit pas si ces précieux fichiers sont accompagnés des métadonnées indispensables, mais il s'agit là, à n'en pas douter, d'une preuve supplémentaire de ce que les archives du XXIème seront électroniques ou ne seront pas...

### **Appel à contribution**

La revue scientifique et technique " Document numérique " prépare un numéro spécial sur le thème : "Archivage et pérennisation , Préserver et exploiter les documents et les données sur le long terme"

#### Contexte

L'archivage électronique – l'archivage numérique - est plus que jamais à l'ordre du jour. Cette actualité est due à plusieurs facteurs :

l'accroissement toujours plus fort de la production numérique à valeur de mémoire pluriannuelle ou pluridécennale : bases de données, sites Internet, publications officielles ou scientifiques, études, actes et contrats, images, correspondance...

l'accompagnement de cette production aux plans juridique (législation sur l'écrit électronique) et technique (structuration XML, norme NF 42-013)

la publication au cours des toutes dernières années d'un panel de textes de références qui façonnent les concepts et la stratégie : norme ISO 15489 sur le Records management, Model requirements for the management of electronic records (Commission européenne), norme OAIS (Open Archival Information system), Stratégie globale pour la conservation à long terme des documents électroniques en Suisse, etc.

le lancement de grands projets d'archivage et de pérennisation sur le long terme, qu'il s'agisse du Web, de l'administration fiscale ou des données techniques des entreprises.

La conservation des données sur le long terme présente un double défi :

identifier les objets de données pertinents, cohérents et indépendants que l'on doit ou que l'on veut conserver, c'est-à-dire des ensembles définis de données qui tracent des actions ou des événements dans un contexte précisé par des métadonnées, constitués de telle manière qu'ils ne soient pas tributaires d'un outil ou d'un système moins durable qu'eux,

pérenniser ces objets dans le temps, en dépit ou à l'aide des évolutions technologiques et sociétales, et les maintenir accessibles aux utilisateurs.

## **Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 13 novembre 2003**

### Objectif

Deux ans après un premier numéro consacré aux débuts de l'archivage électronique, ce nouveau numéro spécial de " Document numérique " souhaite réunir des contributions et témoignages : projets en cours, retours d'expériences, outils et méthodes, présentation de normes et modèles, réflexions prospectives.

Cet appel s'adresse à tous ceux qui sont confrontés à la problématique de l'archivage numérique pérenne : ingénieurs, archivistes et records managers, informaticiens, documentalistes, qualitatifs, juristes, chercheurs.

### Thèmes

Problématique et stratégies d'archivage

L'information numérique et le défi du temps

Normes relatives à l'archivage

Records management et archivage

Les objets d'archivage : documents textes, bases de données, sites Web, courriels, etc.

Modélisation de systèmes d'information intégrant l'archivage numérique

Objets de données et métadonnées

Supports et formats

Questions spécifiques de l'archive électronique : volumes et compression, versionnage

Accès et exploitation des données

Migrations : possibilités et contraintes sur le long terme

Centres d'archives numériques

Le stockage hiérarchique (HSM)

Logiciels libres et pérennisation

Archivage en ligne ou distant

Retours d'expérimentation

Comité de rédaction



**Bulletin des Archives de France sur la conservation à long terme des documents électroniques**  
**N° 13 novembre 2003**

Marie-Anne Chabin, Archive 17, coordinateur  
Françoise Banat-Berger, ministère français de la Justice  
Guylaine Beaudry, Université de Montréal  
Didier Courtaud, CEA  
Marc Fresko, Cornwell Management Consultants, Londres  
Claude Huc, CNES, Toulouse  
Julien Masanès, Bibliothèque nationale de France  
Joël Poivre, Direction des archives de France  
Jean-Daniel Zeller, hôpitaux universitaires de Genève

Calendrier

mi novembre 2003 : diffusion de l'appel  
décembre 2003 : intentions de soumettre un article par l'envoi d'un résumé en français ou en anglais, d'environ 200 mots  
janvier 2004 : réponse aux auteurs  
avril 2004 : date limite remise de la version finale corrigée  
juin 2004 : parution du numéro spécial

L'éditeur et la revue

Revue Document numérique  
Service éditorial Hermès  
LAVOISIER  
14, rue de Provigny  
94236 CACHAN cedex  
Tél. 01 47 40 67 00 - Fax 01 47 40 67 02  
<http://dn.e-revues.com/> ou <http://www.e-revues.lavoisier.fr/>

Recommandations aux auteurs

Les soumissions sont à envoyer à Marie-Anne Chabin (coordonnées ci-dessous) et devront respecter la feuille de style de la revue disponible sur le serveur <http://dn.e-revues.com/> (ou sur demande au coordinateur)

Les articles auront une longueur comprise entre 15 et 20 pages. Ils seront écrits en français ou en anglais. Une bibliographie pertinente et complète est demandée pour chaque article.

Les soumissions peuvent être envoyées sous forme de fichiers PDF (de préférence), Postscript, Word (et RTF) ou papier. Les versions finales seront acceptées sous format Word ou PDF. Dans ce dernier cas (pdf), les corrections ortho-typographiques seront faites par les auteurs.

Les auteurs sont invités à proposer des articles originaux qui n'auraient pas déjà été publiés dans un autre cadre.

Contact

Marie-Anne Chabin, Archive 17  
1, avenue du Père-Lachaise - F-75020 PARIS

**Bulletin des Archives de France sur la conservation à long terme des documents électroniques**  
**N° 13 novembre 2003**

marie-anne.chabin@archive17.fr

prochain Bulletin

Le prochain Bulletin paraîtra en janvier 2004