

LE 15^E CONGRES INTERNATIONAL DES ARCHIVES DE VIENNE (24-26 AOUT 2004)

Louis Faivre d'Arcier

Du 24 au 26 août s'est tenue la partie publique du 15^e congrès organisé par le conseil international des Archives à Vienne, en Autriche. Mon compte rendu ne peut en être que partiel, car le programme, surabondant, prévoyait que quatre voire cinq séances se tenaient en parallèle. La plupart des orateurs que j'ai entendus venaient du secteur public ou para-public ; plus rares étaient ceux qui provenaient du secteur privé, mais je ne peux garantir la valeur statistique de cette remarque (la liste des intervenants qui figure en annexe du programme n'indique pas l'institution à laquelle les personnes appartiennent).

Les thèmes traités dépassent largement la problématique qui se trouvent au centre des travaux du groupe PIN : de nombreuses séances étaient consacrées à des sujets tels que ceux des rapports entre archives et mémoire, archives et société. Cependant, il était possible, au cours de ces trois jours, de ne suivre que des séances sur les rapports entre archivistique et nouvelles technologies, ce qui témoigne sans doute d'une prise de conscience, même si les réalisations concrètes ne sont pas partout tangibles. Je m'en tiendrai ici uniquement aux questions de conservation des archives électroniques.

Plutôt que de rendre compte de chaque séance à laquelle j'ai assisté, je tenterai de dégager quelques-uns des thèmes qui m'ont paru être les plus intéressants : il est toujours possible de se référer au site du congrès (<http://www.wien2004.ica.org>), sur lequel les présentations des différents intervenants devraient être disponibles [c'est du moins ce qui a été annoncé au cours de certaines séances : au 8 septembre 2004, seuls les résumés sont accessibles]. Pour simplifier, parmi les séances dignes d'intérêt, on pouvait entendre soit des retours d'expériences, soit la présentation de recherches en cours ou de textes normatifs.

1. Les retours d'expériences pratiques

Comme je l'ai indiqué, la plupart des intervenants que j'ai entendus paraissaient exercer leur activité dans des services d'archives publics. D'un point de vue géographique, l'origine des orateurs traitant de conservation des documents électroniques ne surprendra pas trop, puisqu'ils venaient soit des pays anglo-saxons (Grande-Bretagne, États-Unis et Australie), soit du nord de l'Europe (Scandinavie, Pays-Bas, Suisse – www.bundesarchiv.ch – et Belgique flamande – www.antwerpen.be/david).

La plupart des présentations laissent apparaître des tendances communes¹.

- **L'historique** des services chargés de la conservation des documents électroniques présente souvent de grandes similitudes. Les Archives ont le plus souvent commencé par les documents statistiques, avant de s'intéresser aux bases de données de gestion et, depuis peu de temps, à d'autres types de documents, devant lesquels les collègues se sentent parfois un peu désarmés. Pour les archives fédérales suisses, le tournant a été pris en 1998 ; en Grande-Bretagne, le changement date de 1997.
- La **quantité** d'informations collectées croît très rapidement. Pour l'instant, aucun service ne parlait d'externalisation de la fonction de stockage, sachant que très peu

¹ On ne peut pas parler de réalisations similaires, dans la mesure où les présentations reflètent parfois des projets en cours ou même, parfois, en passe d'être lancés.

se risquent à une évaluation prospective. Les Archives fédérales suisses proposent les chiffres suivants :

- pour 2005 : 45 km linéaires de papier et 3 To de données avec un accroissement annuel de 2 km linéaires et 2 ou 3 To ;
 - à partir de 2006, et jusqu'en 2015-2020 : accroissement annuel de 2 km linéaires, plus 7 à 22 To.
- Enfin, devant l'augmentation des quantités de documents à traiter, une **automatisation** d'un certain nombre de processus est en préparation dans plusieurs pays, notamment au Royaume-Uni, où les Archives nationales conduisent le projet Seamless Flow, qui intègre tous les aspects automatisables (en l'état actuel de la question) de l'archivage électronique.

L'automatisation peut porter sur :

- le contrôle de la qualité des données versées : une application en service aux Archives nationales du Danemark teste à la fois les formats et la structure des fichiers versés ; c'est également un projet de la ville d'Anvers, et c'est une fonction de l'application SIARD des Archives fédérales suisses, destinée à la préservation des bases de données.
- la récupération automatique des métadonnées, qui constitue l'un des volets de l'application SIARD : il s'agit d'automatiser l'intégration des métadonnées dans le système d'information des Archives ;
- la gestion des supports et des formats : au Danemark, une application permet la migration semi-automatique des formats et des supports des archives versées (semi-automatique, car les documents de la fin des années 1980 posent des problèmes de supports, et ceux de la période 1975-1985 posent des problèmes de formats) ; SIARD ne permet pas de gérer les migrations de formats, en revanche, cette application fait migrer les fichiers de base de données dans un format d'archivage unique.
- la veille technologique (registre PRONOM des formats de fichiers aux National Archives, en Grande-Bretagne).

D'un autre côté, on constate des disparités.

- A certains, l'existence de procédures claires et efficaces de **records management** paraît être un pré-requis à l'archivage électronique : il s'agit des représentants de pays scandinaves, où ces procédures sont réellement en place. D'autres semblent devoir se contenter de plus d'empirisme, lorsque les traditions administratives se rapprochent des nôtres, en attendant de faire évoluer les pratiques (c'est le cas de la ville d'Anvers).
- Le **contexte législatif** diffère selon les pays et oblige les promoteurs des différents projets à introduire des nuances. Au Danemark, la question de l'accès est renvoyée à plus tard ; en Finlande, elle est envisagée immédiatement, car les délais de communication des documents sont plus courts.
Au passage, on regrette un peu l'absence de séance sur les données personnelles, alors que cette question est cruciale pour les Archives.
- Les **moyens** mis en œuvre sont parfois assez conséquents : 15 personnes pour le service chargé des archives électroniques au Danemark, 12 personnes aux National Archives, en Grande-Bretagne. Les Archives nationales anglaises évaluent ainsi leurs coûts, pour les premières années, sachant que des économies d'échelle sont attendues : 28 EUR par entrée d'archives électroniques + 5 EUR/an par Mo stocké.
- Les **stratégies de conservation** diffèrent, ce qui s'explique par l'historique et par les moyens des services :

- la solution adoptée pour la conservation des bases de données est le plus souvent celle des fichiers « à plat », assortie d'une description de la structure originelle ; la conversion au format XML des données elles-mêmes n'est pas toujours envisagée, pour différentes raisons (coût de l'opération et « respect du document ») ;
 - pour les documents bureautiques, le format TIFF a été le plus souvent mis en valeur, car la conservation de la mise en forme – qui fait sens – apparaît fondamentale, mais le PDF/A est parfois évoqué comme une éventualité, et certains conservent les documents dans leur format d'origine (cas du Danemark), du moment qu'il est suffisamment répandu, quitte à opérer des migrations régulières d'une version à une autre du logiciel d'origine ;
 - le XML est utilisé pour structurer les métadonnées automatiquement récupérées, mais pas toujours (car les métadonnées ne sont pas toujours automatiquement récupérées).
- Bien sûr, l'accès par des utilisateurs est considéré comme l'objectif de tout système d'archivage. Néanmoins, aucune solution vraiment simple ne paraît avoir été trouvée. Les Archives du Danemark parlent de la nécessité, pour l'utilisateur des archives, d'avoir une formation initiale en informatique, du moins en l'état actuel de la question et n'offrent qu'une présentation très grossière des données : en vue de l'accès aux bases de données, la structure de celles-ci est convertie en structure relationnelle, et les jeux de caractères ont été migrés (ce qui n'est pas anecdotique, en l'occurrence). En Grande-Bretagne, aucune solution n'a encore été trouvée pour le moment, même si c'est une des priorités du service.

2. Recherches en cours et textes normatifs

a. Une remarquable séance sur les métadonnées s'est tenue le 24 août matin, au cours de laquelle plusieurs initiatives de normalisation des métadonnées ont été présentées.

La **deuxième phase du projet InterPARES** – Interpares2, qui s'intéresse notamment aux questions d'authenticité des documents électroniques – comporte ainsi, un volet relatif aux schémas de métadonnées. Ce registre des schémas de métadonnées se fonde sur un schéma de méta-métadonnées, qui comporte trois niveaux de profondeur. Au total, on compte 120 champs, répartis en 11 zones de premier niveau. Les enjeux d'un tel registre sont de servir de base à l'interopérabilité des métadonnées, en facilitant le mappage des schémas de métadonnées et en permettant d'établir des règles pour l'agrégation des données. Outre le site du congrès, on trouvera plus de détails sur le site d'interpares (www.interpares.org).

Une deuxième présentation était consacrée au projet **Persistent Archives**, qui repose d'abord sur l'élaboration de concepts adaptés à la conservation dans le contexte des nouvelles technologies – et que l'on peut relier à des concepts archivistiques plus classiques (ce que fait Interpares). Il ne s'agit heureusement pas que d'une élaboration conceptuelle : le projet aboutit à tester dans une dizaine de sites pilotes (principalement des universités américaines) un certain nombre d'outils en vue de constituer une « boîte à outils » de l'archiviste. Les outils sont élaborés en référence aux fonctions du modèle OAIS. Le projet pourrait répondre au moins en partie aux préoccupations de services confrontés à de grandes masses de données d'origines diverses, en facilitant à la fois la prise en charge et la gestion des fichiers électroniques. Le site du San Diego Supercomputer Center fournit plus de détails sur le projet : www.sdsc.edu.

Enfin, ont été présentés les travaux relatifs aux **métadonnées du records management**, conduits au sein de l'ISO. Ces travaux ont fait l'objet d'une présentation ici récemment.

b. Par ailleurs, le CIA a terminé un **manuel pratique sur l'archivage électronique**, qui doit paraître d'ici la fin de l'année sous deux formes, papier et électronique. L'approche retenue est avant tout fonctionnelle, l'archiviste se devant de collaborer avec d'autres corps de métier pour bâtir son projet. Le manuel distingue sinon plusieurs fonctions, en tout cas plusieurs profils : celui de l'archiviste, celui de l'informaticien (ce métier se définissant ici par contraste avec celui de l'archiviste) et celui de l'expert en *electronic recordkeeping* (nouveau rôle émergent, mixte d'informaticien et d'archiviste). Il distingue plusieurs étapes de la vie d'un document au cours desquelles l'archiviste doit intervenir (création du *record*, collecte, conservation de l'archive, accès à l'archive). Quiconque s'est occupé de collecte de documents électronique les connaît. L'intérêt de ce nouveau manuel réside à la fois dans une approche « non-linéaire » du problème, au sens où les liens entre les différents stades de la vie d'un document sont montrés ; il soulève aussi les questions qui ont trait à la constitution même des *records* au cours des transactions électroniques. C'est peut-être par ce dernier biais que les archivistes trouveront un ancrage solide pour établir la préservation à long terme des informations numériques.