

Compte rendu du colloque de l'association Aristote : Enjeux et impacts d'XML pour la bureautique

Si, depuis 2 à 3 ans, on parle beaucoup du format d'échange XML, on ne voyait pas encore forcément très bien, hormis pour les extractions à effectuer dans ce format, quelles pouvaient être les implications directes dans le monde du travail : or, aujourd'hui deux suites bureautiques permettent la création, l'enregistrement et le stockage de fichiers en XML. C'est à la faveur de cet événement que l'association Aristote a monté avec l'ADAE un séminaire afin de réfléchir à l'entrée de XML dans le monde de la bureautique : une matinée consacrée à la réflexion sur les enjeux et aux retours d'expérience et l'après-midi où la parole a été donnée à des éditeurs de suite bureautique.

L'introduction a été présentée par Alain Michard spécialiste du format XML depuis de nombreuses années : historique du XML avec d'une part dans les années soixante, des premiers formats pour décrire des pages afin de générer des codes pour les photocomposeuses, d'autre part, le format SGML longtemps confiné au monde technique (aéronautique, énergie...) et enfin, le manifeste en 1994 de Sperberg-McQueen qui présente la nécessité de dépasser les moteurs de recherche du Web. C'est ainsi que XML est apparu en 1998.

XML présente trois exigences " vieilles comme l'informatique " : une indépendance par rapport aux dispositifs physiques, une indépendance par rapport à un constructeur ou à un éditeur particulier et l'universalité des codes.

Aujourd'hui le plus innovant avec XML, c'est son apport dans le travail coopératif sur le Web avec la gestion de contenus en ligne ; c'est également dans le domaine de la dématérialisation et des téléprocédures, la signature numérique incorporées dans la ressource échangée, l'élaboration de formulaires à auto-contrôle ou pré-rempli, les catalogues structurés de services en ligne ; et enfin, dans le domaine de la publication, la possibilité de la publication instantanée (" document dynamique "), son indépendance par rapport au dispositif de présentation, la facilité technique d'agréger des données issues de bases de données.

Alain Michard évoque également la nécessité d'élaborer dans les entreprises avec XML, des gabarits, qui rassemblent ce que l'on met dans les schémas XML, mais également dans les feuilles de style, ainsi que la possibilité de représenter la dynamique des documents (pages comportant des requêtes par exemple).

Enfin, XML est un bon gage de pérennité : archivage de documents classiques figés (comme des images associées à des métadonnées XML EAD, ou encore des formats d'archivage de pages très stable comme eps (encapsuled Postscript)), XML assurant la pérennité des connaissances avec des ressources organisées autour de référentiels communs ; une mémoire associant des données structurées et des documents avec des possibilités de recherche précises.

La stratégie visant à permettre la pérennisation de l'information a été développée par Claude Huc du CNES, qui a rappelé dans un propos liminaire qu'au CNES, des centaines de fichiers Word 2 ont du être re-saisi car ils n'étaient plus lisibles, 7 ans après leur création. Le même problème s'est posé en 2000, pour des fichiers Word 95. Ceci prouve que la problématique de la pérennisation n'intéresse pas seulement le domaine patrimonial mais au-delà, tous ceux qui ont besoin de

Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 14 janvier 2004

préservé l'information au-delà de 5 années (secteurs scientifique, technique, de la santé, de la Défense, du bâtiment...).

Les besoins sont divers : entre pérenniser le contenu et la forme visuelle d'un document, ou bien pérenniser uniquement le contenu et dans tous les cas pérenniser les métadonnées. Le choix est finalement de savoir comment on passe du train de bits à l'information intelligible : soit en passant par un logiciel fermé, dont les spécifications sont secrètes, soit en passant par un logiciel libre ou du moins qui spécifie publiquement ses outils et permet de décrire exhaustivement le train de bits.

Au-delà de cette exigence (documentation ouverte), le choix des formats que l'on fera varie en fonction de l'usage qu'on veut faire du document dans le temps : aptitude à représenter toute l'information, pouvoir modifier le contenu et la présentation, format normalisé ou standardisé, recherche dans le texte... Si l'on veut avant tout préserver l'apparence visuelle, la préférence pourrait aller vers PDF-A ; si l'on veut préserver le contenu uniquement, on choisira plutôt XML.

Les difficultés sont nombreuses : liées aux difficultés de portabilité entre différentes plates-formes, liées au fait qu'un document s'appuie en fait sur un entassement de formats et qu'une difficulté peut apparaître sur telle plate-forme dès lors qu'on retrouve à la base l'utilisation d'un format d'encodage propriétaire (par exemple un jeu de caractère Microsoft Window-1252 à la place du jeu normalisé Iso 8859-1).

Finalement les stratégies possibles sont les suivantes : à partir d'une DTD utilisateur, produire une feuille de style à partir de laquelle on produira un fichier au format de l'outil bureautique utilisé et le faire alors passer par un post processeur pour produire un fichier XML, mais dans ce cas, le post processeur doit évoluer à chaque changement de l'outil bureautique. Alternative plus intéressante pour des documents structurés en très grand nombre : la bureautique est remplacée par des applications métier, celles-ci permettant la saisie (de formulaires, la présentation sur écran, l'édition) et à partir de l'application sont produits d'une part des fichiers XML et d'autre part des fichiers HTML, PDF... L'idéal serait qu'à partir d'une DTD utilisateur, la suite bureautique puisse produire des fichiers XML conformes à la DTD utilisateur et par ailleurs, des fichiers XML conformes à la DTD propre à l'outil bureautique, l'utilisateur ne devant surtout pas voir XML (et ses balises !).

Finalement on voit que les stratégies pour parvenir à XML sont encore incertaines et qu'on en est encore à une bureautique subie et non choisie qui permettrait de pérenniser les documents indépendamment de tout logiciel.

De son côté, Pascal Souhard, de l'ADAE, expose la stratégie de l'agence en matière de répertoires XML. XML est en effet le format d'échanges requis par l'administration dès lors qu'il y a échanges d'informations, à intégrer dans des systèmes d'information. A l'inverse, l'ADAE est méfiante par rapport aux applications propriétaires et incite à l'utilisation des logiciels libres également dans un objectif de constitution dans le court et moyen terme d'un patrimoine numérique.

Actuellement un mouvement est d'ailleurs visible qui consiste pour certains éditeurs à pratiquer une certaine transparence de leurs formats dans leurs nouvelles suites.

L'ADAE est en passe de transformer la position qui était celle de l'ancienne MTIC puis ATICA, d'incitateurs, à celle de prescripteurs : prescriptions de composants réutilisables dans le cadre de gabarits à définir, notamment pour permettre le développement des communications entre

Bulletin des Archives de France sur la conservation à long terme des documents électroniques **N° 14 janvier 2004**

collectivités publiques (A2A).

Le but est de mutualiser les infrastructures en intervenant dans un cadre normatif de mise en oeuvre de nomenclatures et de référentiels. Des groupes de travail sont d'ores et déjà en place (urbanisation des systèmes, sécurité, référentiel de données...) et déjà des répertoires de schémas XML sont publiés à partir des compétences spécifiques de ministères pionniers.

L'ADAE ainsi vise désormais à définir et maintenir des schémas XML. Ainsi, le concept d'adresse : l'adresse n'est pas uniquement géo-postale mais peut également permettre une géo-localisation, comporter des coordonnées x et y pour les travaux d'aménagement du territoire. Il y a autant de modèles d'adresses qu'il y a de communautés d'utilisateurs et d'intérêts. Ce sont ces modélisations répondant à des besoins concrets que l'ADAE va coordonner.

Ainsi le plan stratégique de l'administration électronique a permis de lister 2 à 300 projets qui sont en cours d'études afin de définir des communautés d'intérêts et de mettre en place une méthodologie. A partir de là, seront élaborés par les experts de chaque domaine, des référentiels qui feront l'objet, avant leur publication, d'un appel à commentaires.

Ces travaux permettront la constitution d'un registre des ressources numériques (à partir desquels les éditeurs de logiciels pourront développer leurs nouveaux outils).

A terme, l'objectif est, au-delà de l'interconnexion entre l'intranet ADER et l'annuaire MAIA, le développement d'un portail unique (guichet unique) pour les citoyens qui ont à s'adresser à l'administration dans un souci de simplification : le citoyen, à terme, n'aura à fournir qu'une seule fois un renseignement donné.

Concernant les nouvelles suites bureautiques, Antoine RIZK de la société Valoris présentait les résultats d'une étude menée pour la commission européenne, concernant la stratégie à avoir face à une norme comme XML, l'adaptation des transactions B to B aux échanges entre administrations (A2A), pour les achats publics dématérialisés et avoir une image de l'état du marché concernant les formats de documents ouverts.

Concernant les formats bureautiques, une première constatation : à chaque format, correspond un outil (ainsi pour les suites Microsoft : Word 2000, Word 2001, Word 97, Word 95, Word 6 for Windows, Word 2 pour Windows, Rich Text Format, Word 5.5 pour Dos...). L'interopérabilité totale est quasi impossible.

Des critères ont été établis par Valoris pour juger les formats : critères techniques (ouverture, interopérabilité, pérennité, dynamique du marché, droits : fourniture de la documentation..) et non techniques (non-binaire, modifiable, couverture fonctionnelle, nombre de fichiers, extensible, signature électroniques, intégration d'UNICODE). Or, aucun des formats du marché ne remplit l'ensemble de ces critères.

Valoris a étudié, depuis l'annonce de Microsoft concernant XML, l'offre de Microsoft (la suite Office 2003), la suite bureautique OpenOffice (outils du libre) reprise par le groupe de normalisation OASIS et enfin, la nouvelle annonce de la société Adobe concernant l'association à PDF de XML.

Concernant l'annonce de Microsoft du 17 novembre 2003, le format XML est présent pour le

Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 14 janvier 2004

logiciel de traitement de texte (WordprocessingML), le tableur ainsi que des modèles (formtemplates). Microsoft pour la première fois de son existence, publie une documentation complète et assure sa diffusion gratuite. Les points suivants restent cependant à examiner : les modifications au format, les accords nécessaires à signer (les outils ne sont pas du domaine public : même si les spécifications sont publiques, il faudra passer des contrats avec Microsoft pour éditer de nouveaux outils à partir de leurs sources), les engagements sur de futures évolutions. Quant à la compatibilité avec Open office, elle n'a pas été annoncée.

Concernant Adobe, cette société annonce le nouveau format XML Tagged PDF, qui permet d'avoir dans un seul format, à la fois la structure logique et la structure physique d'un document. Adobe annonce également l'arrivée de formulaires XML combinant une structure logique et une présentation finale. Le client peut alors choisir d'extraire des documents XML pour les intégrer dans une base de données ou d'éditer au format PDF des documents pour des impressions papier. Enfin, Adobe a annoncé un sous-ensemble de Pdf (Pdf-A) format pour l'archivage à long terme des documents, une normalisation à l'ISO pour la fin de l'année 2004 étant attendue. Ceci étant, il n'y a pas de support dans Open office pour le format XML Tagged PDF et on constate également une incompatibilité avec Xforms.

Pour ce qui est OpenOffice, cette suite représente 2% du marché et connaît actuellement un succès incontestable, notamment en raison de sa gratuité (principe du libre). Elle fournit des outils aux couvertures fonctionnelles identiques à celles des suites de Microsoft (ergonomie proche afin que les utilisateurs l'adoptent rapidement et sans douleur), est en open source, permet de réaliser des graphiques, contient un export vers PDF et supporte de nombreuses plateformes (Windows, Linux..). Elle est supportée par Sun, la stratégie étant que Linux passe enfin sur le poste client.

Il reste cependant à examiner précisément quels sont les autres outils qui annoncent une compatibilité avec OpenOffice ; son évolution vers une standardisation (via Oasis), la nature du support des éditeurs, les coûts de migration éventuels. Concernant XML, OpenOffice n'a pas de vision d'une continuité à travers une suite d'outils XML. Elle est très orientée document : il existe une seule DTD par document avec l'impossibilité de créer sa propre DTD. Il s'agit en fait d'un éditeur d'un format donné OpenOffice défini en XML et non d'un éditeur XML. Ceci s'explique également par le fait que OpenOffice a développé depuis plusieurs années la création de documents en XML et n'a pu encore intégrer les dernières avancées dans le domaine (notamment les schémas à la différence de la suite 2003 de Microsoft).

En conclusion, il semblerait qu'à ergonomie comparable, la vision XML et systèmes d'information est plus avancée dans Microsoft, mais se posent alors, concernant cette dernière, tous les problèmes de droits et propriété intellectuelle. Si on tente de se projeter dans les années à venir, il semble que Microsoft gardera une part dominante du marché (à plus de 80%) mais qu'OpenOffice offre une réelle alternative et qu'au final, tous y gagnent à voir apparaître une concurrence plus saine. Deux stratégies sont ici à l'œuvre : Microsoft propose des applications internes pour réaliser des traitements sophistiqués tandis que OpenOffice externalise les traitements, la question étant de savoir jusqu'où on peut aller la sophistication en passant par de traitements externes, qui sont par ailleurs gages d'une meilleure interopérabilité. Quant à l'annonce PDF + XML, il semble davantage représenter un marché orthogonal (et non une concurrence frontale).

Deux retours d'expérience concernant le basculement de Microsoft à OpenOffice sont présentés,

Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 14 janvier 2004

qui concernent tous deux l'administration. Exemple du MINEFI et de son administration centrale, sachant que les services déconcentrés (direction générale des impôts, des douanes) expérimentent à leur tour, ce basculement Le MINEFI s'est lancé dans ce projet dans le cadre du schéma informatique ministériel, autour du groupe de travail normes et standards, sachant que tous les postes bureautiques étaient équipés de Windows 98 et de MS Office 97. L'analyse a été menée en fonction de la politique commerciale et tarifaire de Microsoft d'une part, des nouvelles contraintes budgétaires d'autre part et enfin, de la gratuité d'OpenOffice. Les arguments qui l'ont emporté, sont outre la gratuité d'OpenOffice, son format de fichier ouvert et documenté, sa portabilité, sa pérennisation et son archivage enfin. Par contre, le fait qu'il s'agisse du format XML n'a pas joué, tout autre format remplissant ces conditions aurait également été accepté. En effet, pour les utilisateurs, 95% d'entre eux ont des besoins de bureautique simples, qui ne nécessitent pas XML.

Dans les directions qui ont basculé, les nouveaux postes (500 par an) sont par conséquent livrés avec la suite OpenOffice. Les formations sont rapides : 1 jour pour les utilisateurs standards, 2 à 3 pour les formateurs, le passage d'une suite à l'autre se passant très facilement. En revanche, l'idée de migrer vers un autre système d'exploitation comme Linux a été abandonnée, en raison des problèmes de compatibilité des applications de gestion interfacées avec Microsoft.

Concernant XML, le problème est que pour l'instant, ce format n'a pas été perçu comme un élément d'intégration de la bureautique au système d'information alors qu'il y aurait des besoins dans le domaine de la gestion de contenus, dans l'articulation aisée entre les données et les documents, les échanges facilités avec les applications métier, l'intégration enfin des " cathédrales " Excel construites par certains spécialistes sur leur poste de travail qui n'ont aucune pérennité et visibilité dans le système d'information. Plus généralement, XML ne permet pas de pallier au manque criant d'organisation et méthodes des organismes et institutions.

Bref, il s'agit de savoir ce qu'on va faire du bel outil que constitue XML : un frontal pour accéder aux données de l'entreprise ou, plus ambitieux, un composant du système d'information manipulant un format de données standardisé, indépendant du logiciel de création, permettant l'intégration du poste du travail.

Un autre exemple de basculement vers OpenOffice est donné avec l'académie de Bordeaux, ce basculement concernant un nombre impressionnant d'utilisateurs. Là encore les motivations sont celles du coût d'autant que Microsoft n'acceptait pas de délivrer des licences gratuites pour les enseignants qui préparaient leurs cours et leurs supports chez eux. Là encore le basculement se fait sans difficultés particulières sauf certaines résistances dues au fait que les professeurs notamment de technologie disposaient déjà de supports de cours et d'exercices articulés avec Microsoft, qui ne peuvent plus être ré-utilisés. Des résistances ont pu également apparaître dans certains conseils généraux qui n'acceptaient pas forcément ce choix qui semblait imposé par l'Etat. L'accompagnement en ligne a par ailleurs facilité cette prise en main d'OpenOffice, les quelques difficultés recensées faisant l'objet de réponses précises et circonstanciées.

Le séminaire s'achevait ensuite par la présentation par des éditeurs de nouvelles suites bureautiques. Celles d'Office 2003, d'Open Office et enfin de Koffice, tandis qu'était également présentée la constitution du club des utilisateurs de Star et Open Office , à partir de la communauté OpenOffice francophone, soit l'association CUSSOn (club des utilisateurs de Star et Open office nouveau) qui a comme rôle de faciliter la création et la vie de clubs d'utilisateurs, de récolter des fonds, de lancer des actions de communication et d'animation autour d'Open office et de soutenir des initiatives diverses. Chaque club dispose d'un espace Web et d'une liste de diffusion tandis que l'association

offre un portail d'accès à ces clubs.
Françoise BANAT-BERGER

L'expérience néerlandaise d'évaluation des méthodes de conservation : le cas de l'émulation

Dans le numéro précédent du Bulletin, nous présentions l'expérience originale des Pays-Bas d'un banc d'essai des méthodes de conservation des documents numériques. Le dernier rapport paru porte sur l'évaluation de la méthode de l'émulation.

Il s'agit d'une évaluation globale de la méthode et non d'un retour d'expérience sur un cas concret. L'émulation est définie comme l'utilisation d'un programme informatique (dit "émulateur") sur une plate-forme technique donnée (un ordinateur) qui permette à cette plate-forme de se comporter comme une autre plate-forme technique. Appliquée à la conservation de documents électroniques, l'émulation consiste à :

conserver le document électronique dans le format du logiciel d'origine

conserver également le logiciel d'origine (sous la même forme que le document lui-même, c'est-à-dire un flux d'octets)

écrire un programme d'émulation de la plate-forme technique sur laquelle "tournait" le logiciel d'origine;

lorsqu'on souhaite accéder à nouveau, ultérieurement, au document, on utilise la plate-forme technique du moment sur laquelle on plante l'émulateur. Le logiciel d'origine "tourne" sur cet émulateur et lit le document

Le rapport examine les différentes implications techniques de ce type d'émulation et présente l'état de l'art dans ce domaine. Puis, il discute les avantages et les inconvénients de la méthode.

Au nombre des avantages figure en premier lieu la capacité de l'émulation à conserver l'aspect originel d'un document électronique (puisque l'on utilise le logiciel d'origine pour relire le document) et donc, par la même de garantir son intégrité. Le coût attendu est aussi un argument favorable : en effet, il n'est nécessaire de développer qu'un seul émulateur pour un type de documents indépendamment du nombre de documents à conserver. Les inconvénients, eux, tiennent surtout aux problèmes techniques non encore totalement résolus posés par le principe même de l'émulation.

Pour en savoir plus :

le rapport sur l'émulation sur le site du banc d'essai néerlandais :
http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white_paper_emulatie_EN.pdf

Bulletin des Archives de France sur la conservation à long terme des documents électroniques N° 14 janvier 2004

Le 2 février prochain un Comité interministériel à la réforme de l'Etat (CIRE) consacré exclusivement à l'administration électronique doit se tenir sous la présidence du Premier Ministre. Le Comité doit adopter un plan stratégique pour l'administration électronique (PSAE) préparé par l'Agence pour le développement de l'administration électronique et l'ensemble des Ministères. Ce plan présentera et coordonnera toutes les mesures à prendre pour favoriser l'utilisation des technologies de l'information et de la communication dans l'administration et, notamment, la mise en place de télé-procédures.

Rappelons que le Président de la République, dans ses vœux aux corps constitués a déclaré : " Je souhaite qu'avant la fin de l'année 2006, l'ensemble des démarches [administratives] puissent être accomplies à distance, par téléphone ou par internet."

Le plan comportera également un volet consacré à l'archivage des documents traités ou produits par les procédures dématérialisées.

Signalons enfin, dans le même ordre d'idée, que l'Assemblée nationale vient d'adopter en 2ème lecture le projet de loi pour la confiance dans l'économie numérique, texte qui autorise le vote électronique aux élections professionnelles (élections prud'hommales par exemple).

Pour en savoir plus :

La gazette des communes, n° 1724, 12 janvier 2004, p 38 : E-administration, une percée à confirmer.

Le site de l'adae : www.adae.pm.gouv.fr

Le texte du projet de loi pour la confiance dans l'économie numérique, sur le site de l'assemblée nationale : <http://www.assemblee-nationale.fr/12/projets/pl0528.asp>

A paraître : le manuel sur l'archivage des documents bureautiques

La direction des Archives de France va publier un manuel pratique sur l'archivage des documents bureautiques. Ce manuel fait suite à celui publié en 2002 et qui traitait de l'archivage des documents électroniques en général. L'ouvrage est actuellement sous presse et devrait paraître en mars à la Documentation française. Plus d'informations dans le prochain numéro du Bulletin.

Changement de propriétaire

Le rédacteur actuel du Bulletin, votre serviteur, va prendre, à compter du 1er mars 2004, de

Bulletin des Archives de France sur la conservation à long terme des documents électroniques
N° 14 janvier 2004

nouvelles fonctions au Centre historique des Archives nationales (Minutier central des notaires parisien).

Le prochain numéro du Bulletin sera donc l'œuvre de mon successeur et paraîtra à une date non encore déterminée à l'heure où je rédige ces lignes. Qu'il me soit permis de remercier ici toutes celles et tous ceux qui, par leurs contributions et suggestions, m'ont aidé à faire vivre cette publication.

Joël Poivre