

XML : VERS UN FORMAT DOCUMENTAIRE UNIVERSEL ?

Du nouveau pour le poste de travail

Jean-Pierre ARCHAMBAULT
SCÉRÉN [CNDP] – Mission Veille technologique

Les contraintes qui pèsent sur les budgets informatiques des établissements scolaires invitent à se poser la question du choix des systèmes d'exploitation ou des suites bureautiques : logiciels libres ou logiciels propriétaires ? Au-delà des questions sur le prix des logiciels se pose celle de la pérennité des contenus. Seuls des formats de fichiers publics, comme le format XML, pourraient la garantir. Des choix stratégiques s'imposent.

En octobre 2002, la société Sun a signé avec le ministère de l'Éducation nationale un accord mettant à disposition des établissements scolaires et des écoles sa suite bureautique *StarOffice* à des conditions tarifaires très avantageuses⁽¹⁾. On peut aussi utiliser la version libre de cette suite, *Open Office*, qui fonctionne à la fois sous *Windows*, *MacOS* et *Linux*. Du côté du logiciel libre, il y a également la *Debian Éducation*, issue des travaux d'un groupe de la mission Veille technologique du SCÉRÉN-CNDP (voir encadré page 38) et la distribution *Freeduc* éditée par le CRDP de Bordeaux. Toujours dans le réseau SCÉRÉN, à l'initiative du pôle de compétences logiciels libres, un cédérom multi-plateforme de logiciels pédagogiques libres pour l'école primaire est en cours de réalisation, co-édité par les CRDP de Paris et de Lille.

Il devient donc possible, dans une transition douce et maîtrisée, de diversifier l'environnement informatique des établissements scolaires et des écoles. On constate ainsi des évolutions vers des géométries variables pour les postes de travail : logiciels propriétaires sous systèmes d'exploitation propriétaires ou libres ; logiciels libres sous systèmes d'exploitation propriétaires ou libres ; machines en double amorçage. On peut ainsi conserver des versions anciennes de *Word* pour assurer les échanges de fichiers en tenant compte de l'état exact et actuel de la compatibilité des différents traitements de texte entre eux, et faire coexister sur la même

(1) Voir « Star Office, aperçu avant impressions... », *Médialog* n°45, juin 2003.

machine suites bureautiques propriétaires et libres, les suites libres présentant l'avantage d'être notablement moins coûteuses.

Ces tendances observées sont à rapprocher du pluralisme en train de s'installer pour les infrastructures (serveurs, intranets...) avec les choix faits en faveur de solutions libres dans les académies (*Eole, SLIS, Linux, Apache...*), à l'instar de ce que l'on a enregistré d'une manière générale dans les entreprises et les autres administrations. Le « modèle LAMP » (*Linux, Apache, MySQL, Php*) quant à lui se rencontre de plus en plus fréquemment.

LES BUDGETS INFORMATIQUES

Les évolutions en cours ont à voir avec les budgets informatiques des établissements, des académies, et des collectivités territoriales. On connaît la part qu'y représentent les systèmes d'exploitation et les suites bureautiques. La situation de quasi monopole qui prévaut encore dans ce secteur n'est pas vraiment satisfaisante car elle s'accompagne de tarifs bien trop élevés au regard des usages éducatifs et non commerciaux qui sont faits des logiciels. Le problème ne concerne d'ailleurs pas que l'éducation. En effet, dans un document remis par Microsoft à la SEC (Securities and Exchange Commission), on apprend que l'éditeur a réalisé, au premier trimestre 2002, 86% de marge sur la vente de ses systèmes d'exploitation pour PC (Décision Micro & Réseaux du 25 novembre 2002) : 2,89 milliards de dollars de chiffre d'affaires et 2,48 milliards de dollars de bénéfices. Ce n'est donc pas Office qui lui rapporte le plus, même si la division Information Worker (qui regroupe les logiciels bureautiques) génère dans le même temps 1,88 milliard de dollars de bénéfices pour un chiffre d'affaires de 2,39 milliards de dollars. Quant à elle, La division Serveur Platforms (Windows Serveur, Exchange, SQL) ne rapporte « que » 519 millions de dollars pour un chiffre d'affaires de 1,52 milliard de dollars, les autres unités étant déficitaires.

L'impératif d'une gestion saine et rigoureuse des finances publiques fait que l'on ne peut pas éluder ces questions, dont l'acuité a été renforcée dans la période qui vient de s'écouler par le « nouveau programme » de licences de Microsoft. Ce « nouveau programme » continue à provoquer une émotion certaine dans les académies. Il inquiète également les entreprises et les collectivités locales car il est synonyme de hausse des coûts informatiques.

En supprimant le principe des mises à jour logicielles, qui permettaient par exemple de passer d'une version 5.0 à une version 6.0 sans

La Debian éducation

La distribution *Debian GNU/LINUX* pour l'éducation a été conçue et réalisée dans le contexte général du déploiement des logiciels libres dans le système éducatif. Elle relève des formes de travail coopératif caractéristiques de ces logiciels, et illustre à sa manière ces partenariats d'un type nouveau qui associent créateurs et utilisateurs, service public, collectivités territoriales et sociétés de services. La mission Veille technologique du SCÉRÉN-CNDP a mis en place un groupe de travail, principalement composé d'enseignants qui sont chargés de proposer une offre de logiciels pédagogiques libres, ou fonctionnant dans un environnement libre.

La sélection actuelle⁽¹⁾ comprend un certain nombre de logiciels qui ont été remarqués pour leur intérêt pédagogique, qu'ils concernent directement les élèves ou qu'ils puissent aider les enseignants dans leur travail. Ces logiciels sont accompagnés d'une brève notice descriptive facilitant leur prise en main. Il s'agit d'outils transversaux (traitements de texte, outils Internet), de logiciels d'infographie (retouche d'image, dessin vectoriel) mais aussi de logiciels disciplinaires (géométrie, calcul, calcul formel, visualisation moléculaire, simulation mécanique, schémas électroniques, traitement de textes musicaux, etc.).

La société Logidée a « empaqueté » au format *Debian*⁽²⁾ les logiciels qui ne l'étaient pas, et a mis en place une procédure d'installation extrêmement simple et intuitive, permettant à quelqu'un ne disposant pas de compétences particulières de mener aisément à son terme l'installation de la suite logicielle. Le Conseil général de Seine-et-Marne, en étroite collaboration avec le Centre Départemental de Documentation Pédagogique de Melun, a souhaité mettre à disposition des collèges du département et des enseignants cet ensemble de logiciels. Cette initiative se situe dans le contexte de son « opération 2 000 micros » d'équipement de tous les collèges du département de réseaux avec serveur sous *Linux*, et postes de travail installés en double amorçage (*Linux/Windows*).

(1) Présentation des logiciels et téléchargement : <http://logiciels-libres-cndp.ac-versailles.fr>

(2) *Debian* est un système d'exploitation libre qui utilise le noyau *Linux* et dont la plupart des fonctionnalités proviennent du projet GNU (« GNU's not Unix »).

Le format XML

Au commencement était SGML (*Standard Generalized Markup Language*), norme ISO 8879, langage source à même de décrire des milliers de documents ou de données dans tous les domaines de l'activité humaine. SGML est fondé sur les DTD (*Document Type Definition*, fichier relatif à la définition formalisée d'un certain type de document). Une approche ambitieuse qui n'a d'égale que sa complexité.

HTML (*Hyper Text Markup Language*) est l'une de ces DTD, une version spécifiquement conçue pour le *World Wide Web*, selon le modèle des rapports de bureau, avec titres, paragraphes, illustrations... et quelques fonctionnalités hypertexte et multimédia. HTML, en tant que langage de balisage (*Markup*) est donc davantage concerné par le contenu que par la présentation, conçu de manière aussi indépendante que possible des dispositifs d'affichage. HTML a fait le succès d'Internet.

XML (*Extensible Markup Language*) est une version simplifiée de SGML, optimisée pour le Web, visant à permettre d'échanger des données structurées aussi facilement que les pages HTML aujourd'hui. Une fois que les données sont sur la machine de l'utilisateur, elles peuvent être manipulées, traitées et présentées de manière variée sans nécessiter de retourner vers le serveur (ceux-ci s'en trouvent d'autant allégés). Il permet de s'affranchir du style unique et intangible de HTML sans avoir à s'embarrasser de la complexité de SGML. Il offre une plus grande richesse de présentation et une facilité de recherche améliorée. Il permet également d'exploiter pleinement les metadata, quels que soient leurs formats. Mais surtout, XML

est une architecture, non une application. Un meta langage permettant à des groupes d'utilisateurs de créer leur propre *Markup Language* spécialisé. De cette façon, XML devient une référence irremplaçable dès lors qu'il s'agit de faire communiquer entre eux des systèmes différents.

La nécessité de définir une DTD dans XML demeure toutefois un facteur de complexité. C'est pourquoi a été développé un système plus souple et plus flexible, notamment au niveau du traitement de données non textuelles ou de la possibilité d'établir des liens entre meta-données, dénommé *XML-Schema*.

XML apparaît désormais comme le format privilégié d'échange de données entre applications. À cet titre, il est appelé à jouer un rôle clé de « *lingua franca* », sorte de format pivot ouvrant des passerelles entre les diverses solutions proposées. Face aux enjeux de l'interopérabilité des systèmes, dont l'importance cruciale est de plus en plus ressentie devant les freins résultant des blocages générés par les systèmes propriétaires, XML constitue l'un des facteurs de nature à favoriser l'innovation en organisant et favorisant la liberté des échanges. On le trouve donc fort logiquement au cœur de la vision de la nouvelle génération d'Internet, celle du « Web sémantique » développée par Tim Berners Lee (inventeur du web au CERN en 1989 avec Robert Caillaux) et le W3C.

Alain CHAPTAL

Mission Veille technologique – SCÉRÉN [CNDP]

payer le prix d'une nouvelle licence⁽²⁾, Microsoft fait en définitive payer ses produits beaucoup plus cher. Désormais, il faudra payer une nouvelle licence définitive à chaque fois que l'on souhaitera s'équiper de la dernière version d'un logiciel. À moins d'opter pour la *Software Assurance*, sans que cela constitue véritablement une affaire. En effet, pour adhérer à ce programme – qui permet de faire évoluer son parc logiciel vers les versions les plus récentes de façon automatique –, il faut déboursier 29% d'une licence complète chaque année, soit 116% du prix d'une licence complète tous les quatre ans. Ce prix est à comparer avec celui des mises à jour, qui étaient la plupart du temps deux à quatre fois moins coûteuses qu'une licence complète. On compte de nombreuses entreprises qui adoptent, résignées, ce tarif, pour se sentir rassurées, les produits Microsoft bougeant tellement vite. Si son rythme de renouvellement est inférieur à trois ans et demi environ, un utilisateur a sans aucun doute intérêt

à migrer vers ce nouveau modèle tarifaire. Il y gagnera un peu ou n'y perdra rien, dans le nouveau contexte bien sûr ! Car, par rapport à la période de « l'avant *Software Assurance* » où les mises à jour étaient une pratique courante et beaucoup moins dispendieuse, la facture sera largement supérieure. Pour des cycles de renouvellement supérieurs à quatre ans, il y a tout intérêt à oublier la *Software Assurance* et les établissements scolaires se trouvent généralement dans ce cas. La course à la dernière version d'une suite bureautique n'y est pas la préoccupation essentielle.

À quelque chose malheur est bon, L'épineux problème du support technique a été réglé d'une manière positive. Jusque là, il prenait fin 60 jours après l'apparition de la deuxième grande mise à jour d'un logiciel. Ainsi le support de *Windows 95* a effectivement été arrêté 60 jours après l'arrivée de *Windows 98 OSR2*. Qu'un problème de fiabilité majeur survienne, Microsoft ne mettait plus de correc-

(2) Étude de Gartner dans *JNet solutions* du 11 juin 2002.

tifs à disposition des utilisateurs, quand bien même ils avaient acheté leur logiciel deux ans avant. Comme le faisaient déjà les autres éditeurs, Microsoft propose maintenant une assistance d'une durée minimum de cinq ans, et non plus donc jusqu'à la sortie des nouvelles versions des logiciels. Le directeur technique et sécurité de Microsoft en France admet que « les entreprises n'ayant aucune visibilité sur le planning de sortie de nos logiciels, elles ne pouvaient pas prévoir la durée pendant laquelle elles pouvaient, *a priori*, bénéficier d'une assistance »⁽³⁾. L'organisation de la maintenance des parcs informatiques devraient s'en trouver facilitée d'autant.

LA BUREAUTIQUE XML

Au-delà de l'importante question du prix d'achat des logiciels bureautiques, une suite comme *OpenOffice* présente un avantage de nature stratégique⁽⁴⁾. Le format natif d'enregistrement des documents *OpenOffice* est un format XML public et documenté. Ce qui implique (enfin !) une garantie d'ouverture durable du contenu. Toute application utilisant des outils de développement intégrant les API XML publiques ou des langages de transformation dédiés au XML (ex : XSLT) peut consulter ou modifier les documents sans même utiliser le logiciel *OpenOffice*. Quelle que soit la pérennité commerciale de l'éditeur du logiciel bureautique, l'ouverture des formats de fichiers est la seule véritable garantie de pérennité des contenus. Et même si le format *Microsoft Office* était un standard de fait (ce qui n'est même pas vraiment le cas puisque sa documentation n'est ni publique ni contractuelle, et puisque l'éditeur se réserve le droit de changer de format sans consultation publique préalable à l'occasion d'un changement de version), rien ne vaut, pour les années à venir, le premier format documentaire qui fasse l'objet d'un consensus à peu près universel, à savoir le XML.

Bien que leader historique de la technologie XML en général, Microsoft semble n'adopter cette technologie que tardivement et comme à regret dans l'univers bureautique. La suite *Office* comporte des fonctionnalités d'import-export XML, mais le XML n'est pas encore le format natif dans lequel on peut enregistrer tout document sans aucune perte de contenu, de structure ou de présentation. Pour Microsoft, XML est un format d'échange et non un format de stockage. Cette position s'explique très bien d'un point de vue commercial. En effet, si les fichiers *MS-Office* étaient des fichiers XML, les concurrents commerciaux (Corel, IBM, Sun)

et le monde du logiciel libre auraient tôt fait de délivrer des produits compatibles à 100%, et la compatibilité n'est pas précisément ce qui intéresse les éditeurs en position dominante. Une telle justification n'est évidemment pas très satisfaisante sur le plan technique.

LE PLURALISME

Pour comprendre la logique des formats propriétaires, il faut revenir une dizaine d'années en arrière, en un temps où le traitement de texte et le tableur épuisaient les faibles ressources des machines. En ce temps là, il fallait avoir recours aux algorithmes les plus simples, les plus frustes possibles pour enregistrer ou charger les documents dans des délais raisonnables. L'une des solutions consistait par exemple à copier directement sur le disque les blocs de mémoire physique contenant l'image du document en cours d'édition, et inversement.

Aujourd'hui, cette logique appartient au passé. Les formats documentaires privés n'ont plus de justification avouable. Aucun argument technique sérieux ne vient étayer la théorie qui oppose formats de stockage et formats d'échange. De nos jours, l'abandon des formats de fichiers bureautiques propriétaires au profit de formats publics est devenu à la fois stratégiquement nécessaire pour les utilisateurs et techniquement facile pour les éditeurs.

Le pluralisme est une valeur qui, pour le domaine technologique comme pour les autres, fait l'objet d'un consensus dans la communauté éducative. Il est en phase avec la culture enseignante et les missions du système éducatif. Il reste cependant encore un objectif à atteindre pour les environnements informatiques des établissements scolaires et des écoles, essentiellement pour le poste de travail. Les évolutions récentes l'ont placé à portée de main. Les budgets informatiques ont tout à y gagner. Les démarches pédagogiques de coopération et de mutualisation aussi, car il signifie des logiques d'ouverture (des standards et des logiciels) qui permettent pleinement la compatibilité et l'interopérabilité sans lesquelles il est vain d'espérer pouvoir travailler en commun. ■

(3) *Décision Micro & Réseaux* du 21 octobre 2002.

(4) Voir l'article de Jean-Marie Gouarné « OpenOffice.org : vrais enjeux et idées fausses », *JDNet solutions*, 15 janvier 2003. http://solutions.journaldunet.com/0301/030115_trib_gouarne.shtml