

**Informatique & Bible, asbl - Belgique**  
**Rue de Maredsous, 11 B5537 Denée - Belgique**  
**Tél:+32(0)82.69.96.47 Fax:+32(0)82.22.32.69**  
**cib@cibmaredsous.be**



Interface n° e-117 Décembre 2009

[Informatique et patrimoine: d'une collection en péril à un réseau de collaboration](#)

[Séminaire organisé par La Maison de la Métallurgie et de l'Industrie de Liège & Informatique & Bible, asbl , Maredsous, 29 octobre 2009](#)

[Contribution](#)

[Méthodologie intellectuelle, matérielle et logicielle](#)

Plusieurs institutions collectent et mettent à disposition à travers Internet des "archives" plus ou moins considérables. (jusqu'à 3 petaoctets pour archive.org) de textes, de sons, vidéos et même de logiciels. Ces "archives" semblent être plus de l'ordre du stockage accumulatif de données diverses en leur état au moment de leur entrée dans ces archives. Faut-il se contenter de ce type d'accumulations ou doit-on travailler à un autre mode de conservation à moyen et long terme dans le secteur des patrimoines informatiques?

**QUI** : la préservation d'un patrimoine ne doit pas être réservée aux spécialistes de l'archivage mais devrait concerner chaque créateur de données, y compris le particulier à qui l'on impose par exemple les documents administratifs électroniques et qui est donc responsable de leur conservation à plus ou moins long terme suivant des règles légales.

**QUE, QUOI** : étant donné la croissance exponentielle du nombre d'informations électroniques créées chaque jour par chacun, la première démarche consiste à définir CE qu'il faut conserver, le type de données, et ceci à chaque niveau de création, du particulier, aux administrations en passant par les PME, PMI.

**DONT** : quels états des données faut-il conserver? Si on tente, par facilité, de tout conserver, on atteindra vite des volumes de données non maîtrisables, même si les capacités de mémoire ne cessent d'augmenter avec des encombrements toujours moindres, mais avec des coûts de gestion importants.

**OÙ** faut-il conserver ces données? faut-il se fier à des mémorisations décentralisées sur quelques gros serveurs détenus par des sociétés privées? Quelles obligations de conservation ont ces sociétés au-delà du minimum légal? Sait-on toujours où se trouvent ces données, dans quel pays? quel continent? Et quid en cas de défaillance ou de faillite d'un tel opérateur?

**COMMENT** conserver ce qui doit l'être? Il faut distinguer plusieurs types de données:

- Pour les données "imprimables" (textes, photos) une norme d'archivage, PDF/A (ISO 19005) tente de s'imposer. Si l'idée de base est bonne: indépendance, autonomie..., elle impose certaines restrictions (non inclusion d'objet dynamique, interdiction du lancement de code script ou exécutable, inclusion de toutes les polices de caractères, UNICODE, interdiction de chiffrement et de sécurité...) et son application est questionnable. Si chacun connaît et produit des PDF, combien de ces fichiers sont-ils compatibles avec la norme PDF/A? Expérience faite, presque aucun document PDF n'est directement compatible PDF/A et certains PDF/A ne passent même pas le test de validation avec succès. De plus cette norme ne s'applique aujourd'hui qu'à des documents "imprimables", donc des textes et des images. Sont exclues les données conservées en "base de données" si elles ne sont pas exportées sous formes d'états imprimés, les sons, les vidéos....

L'archivage PDF/A est donc une démarche volontaire, complexe et coûteuse dans la mesure

où les logiciels de validation proposés sont eux-mêmes coûteux. Olaf Drümmer affirme avec un optimisme certain, que cette norme s'imposera le jour où la masse critique sera atteinte, ce qu'il évalue à 2010!

- Les données non imprimables, ou non imprimées (fichiers de données, bases de données...), doivent être conservées dans un format indépendant de tout logiciel. Le format aujourd'hui reconnu est l'XML en UNICODE, accompagné de sa structure (DTD ou schémas) et de sa mise en forme (CSS...).

- Les données interactives? tels les sites WEB? aucune directive ne les concerne. Une initiative américaine archive.org archive TOUTES les versions de TOUS les sites web. Mais là encore, la démarche est questionnable de plusieurs points de vue, aussi bien des droits d'auteur (chacun a le droit d'exclure son site de cette procédure robotisée, par une démarche opt-out!) que de la pérennité (quelles obligations dans la durée?).

- Les logiciels: Là encore la nuance s'impose: Les logiciels propriétaires, ceux que l'on a créé et dont on possède les "sources" en langage clair, lisible et donc "imprimable" peuvent être conservés mais pas indépendamment de l'environnement sur lequel ils "tournaient". Les programmes ou logiciels que l'on a simplement utilisés et dont on ne possède pas les sources pourraient être conservés avec leur environnement d'exécution.

La conservation de ces logiciels ne peut se concevoir qu'en lien avec le matériel sur lequel ils tournaient.

Au-delà de ces normes (PDF/A, XML...) il ne faut pas oublier la question des supports. Tout support électronique doit être vérifié périodiquement quant à sa fiabilité de lecture, le slogan "avec P..... vous gravez pour l'éternité" n'est que publicité mensongère. Seule la migration des supports garantit leur lisibilité future. Mais dans ce cas, seules les données sont conservées comme patrimoine. Les supports eux-mêmes ne peuvent être conservés QUE comme une trace matérielle de l'évolution de ces techniques, et dans ce cas, un seul spécimen suffit.

CEPENDANT , l'idéal n'étant jamais atteint, la conservation, même non normalisée ni organisée, ne vaut-elle pas mieux que l'oubli? À charge aux historiens de faire revivre ces patrimoines!

NÉANMOINS , chacun peut déjà apporter sa contribution à une meilleure conservation en appliquant des règles simples de nom de fichier, de copie de sécurité et en utilisant, au jour le jour, les outils les plus adaptés à une certaine indépendance et longévité (odt plutôt que doc, xml plutôt que mdb ou dba...).

TOUTEFOIS la première question n'est-elle pas: "le monde virtuel qui nous envahit n'appellera-t-il pas une conservation virtuelle?". Ou encore: "comment adapter le principe de conservation au monde virtuel alors que la conception même de la conservation reste très matérielle?".

Ces réflexions constituent la modeste contribution d'informaticiens qui, depuis plus de 30 ans, sont passés des main-frames aux PC, des cartes perforées aux clés USB en passant par les disquettes de 8", 5" et 3", les bandes magnétiques, les disques optiques, les cd et les dvd.

Tout cela au prix de nombreuses migrations mais sans aucune perte de données et en conservant toutes les données, exploitables aujourd'hui, même si ce n'est pas dans leur environnement d'origine.

