

Les coûts d'une conservation électronique

Alors que la conservation de documents papier consomme manifestement de la place dans les dépôts, la conservation de données électronique semble être illimitée avec les capacités de stockage de nos ordinateurs qui augmentent à des vitesses vertigineuses. Et pourtant, la conservation électronique a des coûts que l'on ignore souvent (ou feint d'ignorer). Le point avec un informaticien.

De la gestion de dossiers papier ...

La question de la gestion de la vie des documents a toujours été de celles que l'on écarte: tant qu'il reste de la place au sous-sol, on entasse les classeurs fédé-raux d'archives ...

S'il est vrai que cette pratique a prévalu durant des décennies, s'il est vrai aussi qu'elle a permis par la suite à des chercheurs de retrouver des informations extrêmement intéressantes, toutes les entreprises ont dû se résoudre à mettre en place une gestion de la fin de vie des documents papier et prévoir la destruction d'une grande partie d'entre eux. Les dépôts en effet sont difficilement extensibles et tous les classeurs n'ont pas la même valeur au cours du temps.

... à la conservation de données électroniques

Or, à l'heure de la création des documents informatiques tout azimuth, la place occupée sur des supports électroniques par les documents et les archives est globalement très faible. Passer du simple au double, au décuple même, de données ne change pas fondamentalement la donne: on a suffisamment de place à la cave pour quelques cassettes d'archives ou quelques disques durs.

Pour un volume donné, la densité de stockage augmente chaque année. On peut donc admettre que tout conserver n'occupera jamais davantage de place: à volume égal, un simple disque dur contenait 1 Go il y a vingt ans, et 10 To (soit 10 000 Go) aujourd'hui! Les cassettes d'archive doublent leur capacité tous les deux ans environ. À elle seule, cette augmentation des capacités de stockage absorbe aisément, année après année, la somme totale des documents à conserver, anciens et nouveaux. Pour les documents bureautiques en tout cas ...

La quantité de données à conserver augmente année après année, mais la densité de stockage augmente encore plus vite. Le volume physique occupé par les archives n'augmente donc pas.

La question est donc évidente: pourquoi se préoccuper de ces données, si le coût induit par leur stockage physique n'augmente pas?

Pour tenter de répondre à cette question, il faut pouvoir la cerner.

Tout d'abord, il convient de différencier l'archive papier, dont la principale caractéristique est d'être directement lisible, de l'archive informatique qui a besoin d'un support et d'un outil pour être lue.

Formats et supports des documents électroniques

L'outil nécessaire pour lire un document électronique, c'est le logiciel avec lequel on a créé le dit document. Un document informatique sous son format d'origine – Microsoft Excel, Auto-CAD, etc. – ne pourra bien souvent plus être lu après quelques années de stockage: le format a évolué à un tel point que le logiciel actuel ne pourra tout simplement plus l'ouvrir. Les filtres d'importation vont gérer quelques versions antérieures, mais généralement pas les versions très anciennes. Il arrive également qu'un logiciel disparaisse du marché (Works par exemple). Faute d'outil, les documents créés dans ces formats sont devenus illisibles, ou très imparfaitement gérés par les filtres d'importation.

Le support, c'est la partie visible de la conservation des documents électroniques: disque, cassette, clé USB, etc. Tous les supports informatiques sont fragiles et sont sujets au vieillissement. Qu'ils soient optiques ou magnétiques, ils se dégradent inévitablement. Durant un certain temps, les redondances de données intégrées au support lui-même permettent de pallier cette dégradation en recomposant les portions perdues du message, mais finalement le contenu devient illisible, il est perdu. De plus, les supports statiques (clé USB, disque SSD) peuvent également tomber en panne. La difficulté réside dans le fait que, visuellement, rien n'indique que le support se dégrade.

Contraintes des supports informatiques

Certaines catégories de supports sont dévolues à l'archivage et sont certifiées pour rester lisibles durant des décennies [1](#). Toutefois, devant l'impossibilité de vérifier cet argument, dans les milieux de conservation des données, on recommande la recopie intégrale des documents tous les cinq ans au moins, et au minimum sur deux supports indépendants.

Une fois perdu, un support peut encore être sauvé par des entreprises spécialisées: la récupération en salle blanche est généralement possible, mais à un coût très élevé. Enfin, afin de parer la destruction par le feu, il convient de stocker des copies en plusieurs lieux.

Rappelons-nous finalement qu'il faudra encore posséder un lecteur en état de marche, capable de lire le disque ou la cassette. Qui possède encore un lecteur de disquettes? C'était pourtant le format standard il y a vingt ans. Combien d'archives comptables dorment sur ce support?

Alors, quel peut être le sens de conserver de tels documents sous cette forme, quand bien même leur stockage serait bon marché? On croit posséder des archives, mais on possède en fait de la poussière, en tout point comparable à un classeur dévoré par les souris.

Solution à la préservation électronique?

Existe-t-il une solution pour préserver ces données volatiles? Oui, elle est composée de plusieurs éléments indissociables, que nous avons vus précédemment:

- Éviter pour l'archivage les formats de données propriétaires ou non pérennes, leur préférer les formats d'archives reconnus et standardisés: PDF archive, images standards, bases de données non verrouillée [2](#)
- Renouveler périodiquement les supports. Cela permet de profiter de l'augmentation de densité des dernières générations de produits, mais il faudra tout remplacer: un ancien lecteur de bandes ne tirera aucun profit des bandes à haute capacité. Il va donc falloir remplacer le lecteur ou le robot de gestion des cassettes. Il en va de même pour les disques durs.
- Contrôler le contenu des supports: les systèmes de stockage et d'archivage doivent périodiquement lire et comparer les données enregistrées, et ce automatiquement, afin de vérifier qu'elles sont encore lisibles.
- S'assurer de disposer d'une copie de sécurité, toujours située à l'extérieur du bâtiment principal, indispensable pour gérer les problématiques d'inondations, de feu et de vol.

Le stockage brut de données informatiques est donc totalement inutile si l'on souhaite pouvoir relire les données à l'avenir.

Gérer la masse des documents et des données

L'augmentation de la densité des capacités de stockage permet d'absorber au fur et à mesure les nouvelles données, pour autant qu'il s'agisse de données de bureautique dont la quantité augmente linéairement chaque année. Or il se trouve que ce type de données représente une part de plus en plus congrue des documents essentiels au fonctionnement de l'entreprise.

À l'inverse, les documents de travail des systèmes métier actuels représentent de très loin la plus grosse part de l'information d'une entreprise et leur masse augmente de manière extrêmement rapide et massive: dossiers systématiquement scannés et stockés dans une application, prises de vue par drone, relevés topographiques, rendus 3D, photographies en haute résolution, films, radiographies médicales, IRM, scanners de terrain, scanners 360°, etc. Ces données, pour un simple dossier client, peuvent aisément représenter plusieurs dizaines de gigaoctets de données, voire bien davantage.

L'augmentation de la densité du stockage des données n'est de loin pas aussi rapide et les systèmes de stockage à plusieurs disques durs sont devenus la norme.

Augmentation des données plus grande que la baisse des prix

Prenons pour exemple un bureau technique de 20 personnes. Aujourd'hui, ce bureau aura à gérer environ 15 à 20 To de données. Le stockage d'une telle quantité d'information et leur réplication dans un centre de données (ou datacenter) sécurisé en Suisse va représenter à lui seul un coût minimum de CHF 1500.– par mois.

Certes, ce coût va baisser, probablement de moitié tous les trois ans [3](#), mais en revanche, durant ce même laps de temps, la quantité de données à stocker va quintupler [4](#)!

La raison de cette augmentation est simple, les appareils qui créent ces données deviennent de plus en plus gourmands en stockage: photos et films à résolution en hausse constante, utilisation systématique des relevés par drone et des scanners 3D, systématisation de la numérisation de documents (dossiers de candidature, courrier, factures), etc.

Le résultat est que la masse de données à stocker croît de manière bien plus rapide que la baisse de prix des espaces de stockage. Cette dernière est relativement linéaire, alors que l'augmentation de la quantité de données a tendance à suivre une courbe exponentielle.

La question de l'archivage de ces données devient donc très rapidement une nécessité économique: il n'est pas du tout raisonnable ni viable à long terme de conserver toutes ces données, ce d'autant plus que la majorité de ces données ne sera bientôt plus lisible si aucune politique de préservation n'est mise en œuvre. Ces données deviendront donc parfaitement inutiles, mais leur stockage continuera de coûter!

Ainsi, sans prendre de mesures de gestion de la vie des documents, cette entreprise va donc dépenser des fortunes pour stocker ... du vent. Si elle veut proposer des prestations meilleur marché et rester concurrentielle, elle doit impérativement maîtriser ce type de charges!

Gérer la vie des documents dès maintenant pour maîtriser les charges

Que faire alors? Un bon plan d'archivage repose sur une gestion de la vie des documents! Et c'est un informaticien qui le martèle!

Il est indispensable de penser sur le long terme et de prévoir une organisation des documents qui permette leur archivage intelligent: par exemple le passage d'un dossier actif de client en mode archive figée qui nécessite moins de ressources de stockage (car moins de duplication des données), la suppression des documents devenus inutiles dans le cadre de tel contrat, la conservation des documents essentiels sous un format PDF/Archives stable ou encore la conservation d'un minimum de sources brutes.

Nous sommes conscients que la gestion de cette problématique n'est pas aisée, elle est de surcroît chronophage. Se faire aider d'un professionnel en gestion documentaire allège singulièrement ce travail et évite les écueils habituels. C'est maintenant qu'il faut y penser, pas dans dix ans!

1 Parmi ces tentatives de conception de supports voulus comme permanents, mentionnons le support sur verre développé par l'entreprise japonaise Hitachi: Schwan, Ben, «Speicher aus Glas», in: Technology Review, 26.10.2012, <http://www.heise.de/tr/artikel/Speicher-aus-Glas-1729182.html> Plus récemment, un projet de l'École polytechnique fédérale de Zurich (ETH) cherche à utiliser de l'ADN comme support d'archivage pérenne: «ETH-Forscher machen DNA zum Langzeit-Datenarchiv», in: ICT kommunikation, 16.02.2015, <http://ictk.ch/content/eth-forscher-machen-dna-zum-lang-zeit-datenarchiv>

2 On trouvera un «Catalogue de formats de données d'archivage (Cfa)» sur le site du Centre de coordination pour l'archivage à long terme de documents électroniques (CECO), une entreprise commune des Archives suisses: <http://kost-ceco.ch/wiki/whelp/Cfa/>

3 C'est en tout cas le rythme constaté jusqu'à maintenant.

4 Là encore, c'est l'évolution passée qui indique cette tendance.



Luc Schneider Tourn

Luc Schneider Tourn est co-fondateur de Practeo SA, société active depuis près de 15 ans dans le soutien et l'accompagnement des PME romandes dans les différents domaines touchant à l'informatique: gestion d'infrastructure, archivage des données, développement d'applications spécifiques et communication web.

Résumé

Deutsch

Die Kosten der elektronischen Aufbewahrung

Während die Erhaltung von Papierunterlagen offensichtlichen Platz in Magazinen beansprucht, scheinen die Möglichkeiten zur Aufbewahrung und Erhaltung von digitalen Daten mit den Speichermöglichkeiten unserer Computer, die rasant zunehmen, unbegrenzt zu sein. Und dennoch, die Erhaltung digitaler Daten verursacht Kosten, die häufig ignoriert werden (oder bei denen man so tut, als ob man sie nicht kennt). Die Speicher und die Formate für die Erhaltung sind nicht ewig haltbar, ebenso wenig wie die Lesegeräte: Man muss deshalb Entwicklungen und Veränderungen verfolgen und die digitalen Daten regelmässig auf neue Speicher und in neue Formate migrieren. Zusätzlich gilt, dass die Datenmenge rascher zunimmt als die Preisabnahme für den dafür benötigten Speicher. Es kommt deshalb teurer, alles aufbewahren zu wollen. Vielmehr sollte man diejenigen ausgewählten Daten bewältigen können, die man erhalten möchte, und dies ab ihrem Entstehungszeitpunkt.