



La conservation des données

Dans de nombreuses réunions techniques, certains thèmes ou processus métiers font l'objet de discussions animées, voire de querelles sémantiques. La *conservation des données* est l'un des sujets favoris de la période actuelle. Dès que ce thème est évoqué, les commentaires pour l'expliquer affluent ; on parle de stockage, de sauvegarde, d'archivage, de continuité de service et de bien d'autres domaines. Les passionnés vont jusqu'à évoquer le records management, les e-mails, l'ILM¹, la GED², le HSM³ et les migrations.

Mais ce sujet est-il réellement complexe ou marque-t-il la confusion des domaines interpellés?

La première question à se poser concerne l'*objectif* de cette conservation. Voulons-nous conserver les données pour recours en cas d'incident ? Dans ce cas, nous abordons le domaine de la *sauvegarde*, puis de la reprise après sinistre ou même de la continuité de services. Rappelons que les données sauvegardées sont remplacées périodiquement par celles des nouvelles sauvegardes, et que la restauration est une fonction primordiale qui doit être testée de façon régulière.

L'*archivage*, de façon très différente, vise à conserver les données d'origine pour des accès ultérieurs, souvent pour de longues durées. Cette conservation des données est organisée pour les besoins opérationnels de *consultation* des entreprises ou pour répondre aux contraintes de *conformité* aux lois et réglementations. Notons que les données sont stables : elles ne doivent subir aucune modification après archivage. Très proche, le *records management* s'applique aux données non modifiables, est identique à l'archivage mais ne comprend pas les archives historiques.

Il serait possible de montrer que chacun des thèmes précédents présente des caractéristiques qui le distinguent des autres. Ainsi la *GED*, dédiée à la gestion quotidienne des documents, autorise la modification des documents et autorise leur destruction par leur auteur. Elle ne correspond pas à l'archivage. La *HSM* est une technologie qui facilite l'automatisation d'une stratégie de stockage multicouche mais qui n'est utilisée que dans certains types d'archivage.

Introduire l'*ILM* permet d'élargir le débat. En effet, l'ILM est une approche pour évaluer puis gérer l'information en fonction de sa valeur pour l'entreprise au fil du temps. Elle n'est pas une technologie ni une solution, n'est pas un produit matériel ni logiciel, et ne se réduit pas à l'archivage ni à la gestion de contenu. Elle comprend toutes les fonctions qui composent le cycle de vie des informations, de leur création jusqu'à leur destruction.

Conserver les documents nécessite des systèmes appropriés. Ainsi, les qualités requises pour les archives – inaltérabilité, fiabilité, résilience aux changements, disponibilité, etc. – impliquent que le système possède des attributs fonctionnels adaptés. Le scellement est garanti par les technologies *WORM*⁴ (physiques ou logiques), l'authentification et l'intégrité sont assurées par les *coffres-forts électroniques*, l'accès au contenu est couvert par les systèmes *CAS*⁵, la redondance et l'évolutivité

¹ Information Lifecycle Management : Gestion du cycle de vie de l'information

² Gestion Electronique de Documents

³ Hierarchical Storage Management : Gestion hiérarchique du stockage

⁴ WORM : Write Once Read Many

⁵ CAS (Content Access System) : Système d'accès par le contenu



sans arrêt de service sont consolidées par la technologie *RAIN*⁶, et la maîtrise du TCO dépend du stockage *multi-technologie*.

Mais l'accumulation d'années de conservation pose des problèmes aigus : il faut que les archives, par exemple, résistent à des *changements technologiques* (système, application, format ou support) ou de comportement des utilisateurs. Bien qu'aucune technologie de base (disques magnétiques ou optiques, bandes) ne soit imposée, l'archive doit être non réinscriptible, non effaçable et restituable. Quel que soit le système choisi, il est nécessaire, actuellement, de mettre en place une stratégie de *migration des données* et de provisionner les budgets nécessaires. Le choix et l'utilisation de systèmes de stockage (SAN) capables d'absorber les mouvements de données complètent ces exigences.

D'autres thèmes pourraient enrichir l'intérêt et les contraintes de la conservation des données. S'intéresser à la gestion de contenu, à la conformité aux lois et réglementations, aux aspects économiques, aux fournisseurs et prestataires de services, aux modèles et méthodes, et aux meilleures pratiques... pourront faire l'objet d'autres analyses.

⁶ RAIN (Redondant Array of Independent Nodes) : Système redondant de noeuds indépendants