PAS@PAS

Archiver, conserver.

Conserver des produits, des ressources ou des données au-delà de leur durée d'utilisation justifiée par les besoins principaux à l'origine afin de satisfaire à des besoins secondaires et hypothétiques.

Règle d'or.

Les données sont une richesse.

Quelles puissent être valorisées ou quelles engendrent des coûts importants en cas de perte, les données font partie du patrimoine de l'organisation.

Comme toute richesse, il faut :

- Identifier celles qui ont de la valeur (juridique, économique, historique),
- Les collecter en garantissant le niveau de qualité attendu,
- Les conserver à l'abri des menaces et de l'usure du temps,
- En garantir l'accès en cas de besoins,
- Vérifier régulièrement l'état de ce patrimoine.

Tout n'a pas de la valeur. Il vaut mieux constituer un patrimoine de données limité mais dont on est assuré de pouvoir disposer en tant que de besoin.

L'acte de conservation (pas uniquement d'archivage d'ailleurs) est un acte particulièrement important. L'information est la sève de l'entreprise.

A ne pas faire.

Tout garder.

Dans le doute, il peut être tentant de tout conserver, de tout archiver. Le problème est que l'archivage a un coût de mise en place, d'alimentation et de conservation. Conserver des données qui ne seront jamais utilisées est un gaspillage.

L'archivage doit être un acte raisonné. Il faut s'interroger sur les raisons qui motivent la conservation. Elles permettront de déterminer le périmètre, la forme et la durée de conservation. Y compris les approches de type bigdata s'avèrent bien plus performantes lorsqu'elles reposent sur des données choisies et collectées dans un but précis que lorsqu'elles tapent dans un amas de données dont on ne maîtrise pas la qualité.

Toutes les archives dont la raison d'être n'est plus valide doivent être détruites.

Ne rien archiver.

Certaines organisations n'archivent pas mais conservent toutes les données en données vivantes.

Le problème est que :

- Le coût de traitement des données vivantes est bien supérieur à celui des archives.
- L'accumulation de données pèse rapidement sur les performances.
- En conservant tout, on évite de se poser la question du pourquoi. Quand vient l'obligation de purger, il n'est pas rare que la suppression se fasse sans discernement. Il est toujours plus complexe et coûteux de trier un amas de données à postériori que de le faire au fil de l'eau.

A l'autre extrême, certaines organisations ne conservent pas les données. Elles les suppriment dès lors qu'elles ne sont plus vivantes. Ne jamais s'assurer de la qualité de ses archives ou ne pas tester les procédures d'accès et de restauration revient exactement au même. Ce faisant, l'organisation s'expose à :

- Ne pas pouvoir tirer bénéfice de certaines informations de valeur,
- Ne pas pouvoir faire la preuve d'actes passés,
- Voir même, être dans l'obligation de reconstituer des données perdues.

Les risques.

Surcharger inutilement les systèmes.

Les causes principales sont :

- Un manque de réflexion et de discernement dans le choix des données à utiliser et à conserver. Même si le stockage physique est de moins en moins cher, les techniques de stockage sont gourmandes.
- Conserver des données qui ne seront jamais utilisées ou qu'il est très aisé de reconstituer. Dans le premier cas, à quoi cela sert de conserver les données qui ont servi à l'élaboration d'un document si la conservation du document suffit. Dans le second cas, à quoi cela sert-il de conserver des données qui sont également disponibles ailleurs.

Être dans l'impossibilité de faire face à des obligations.

Les causes principales sont :

- Une absence de réflexion et de veille sur les données à conserver et la durée de cette conservation. Les principaux motifs de conservation sont légaux ou économiques. A minima pendant la validité de l'activité à son origine. Parfois, cette durée peut être très longue. Par exemple une centrale nucléaire a une durée de vie supérieure à 40 ans. Les données sociales liées à la retraite sont valides pendant toute la vie de la personne concernée.
  Des durées de conservation courtes peuvent être imposées pour des raisons de protection des données personnelles (le droit à l'oubli par exemple).
  Sur une même donnée, des exigences peuvent être contradictoires, la prudence veut de se caler sur celle la plus durement sanctionnée.
- Des données mal conservées ou de mauvaise qualité soit dès l'origine, soit à la suite d'une altération. En cas de besoin, non seulement il peut être nécessaire de produire les données mais également d'apporter la preuve qu'elles n'ont pas été altérées lors de leur captation et de leur conservation.
  La qualité des données, la sécurité et la sureté de la conservation et de leur restitution sont tout aussi importantes que le fait de posséder la donnée elle-même.

Ne pas avoir accès aux données conservées.

Les causes principales sont :

- Ne jamais avoir vérifié les procédures d'accès et de restauration. Sauvegarder ou archiver des données sans vérifier régulièrement le processus de restauration ou d'accès ne sert strictement à rien.
  Simuler une perte ou une destruction des données n'est jamais une perte de temps. Cela permet de se rendre compte que la restauration est bien plus complexe qu'elle ne le semble. La restauration d'un site prend généralement beaucoup plus de temps qu'annoncé.
- Avoir perdu les données faute de réplication et protection suffisante. Par excès de confiance dans les infrastructures physiques de sécurité des centres de calcul, par économie souvent, une seule copie est réalisée. Le plus souvent, elle est stockée sur le site principal.
  La barrière du coût en matière de redondance de sites vient souvent du fait qu'on cherche à préserver toutes les données sans discernement. Répliquer sur un site distant les données réellement critiques constitue toujours une bonne assurance.
- Ne pas avoir actualisé le format des données ou les infrastructures pour les gérer. Surtout dans le cas des données conservées sur des périodes longues, on ne cherche que rarement à les changer de format et de support pour éviter de devoir faire face à une

obsolescence technologique. La traditionnelle matérialisation sous forme de microfiches reste une option parfaitement valide et peu coûteuse.

Les pistes de solutions.

Identifier les besoins de conservation.

Par analyse des objectifs et des processus, identifier les données essentielles au fonctionnement de l'organisation. Par une veille technique et réglementaire, identifier les données à conserver (ou ne pas conserver) pour des raisons légales. Par une analyse des tendances du marché, identifier les données manipulées par l'organisation qui peuvent faire l'objet d'une valorisation.

Identifier les durées et les stades de conservation. Le stade de conservation correspond au fait de transformer la donnée à un certain moment. Par exemple, certaines données publiques ont une durée de conservation légale. Avant destruction, un échantillon est transféré aux archives nationales. Dans d'autres cas, ce peut être que le périmètre de données soit réduit ou que la forme de conservation change.

La seule raison « historique », « au cas où » ne constitue pas une raison valable.

Hiérarchiser les données par niveaux de criticité. Trois niveaux suffisent généralement :

- Non essentielle ou reconstituable. A bien y regarder, c'est souvent la catégorie dominante car on dispose souvent des mêmes informations ou documents en plusieurs exemplaires. Dans de nombreux cas, on peut également demander à un tiers de les refournir.
- Essentielles et onéreuses à reconstituer. Ce sont les données dont l'absence, même temporaire peut générer des coûts de reconstitution ou de mise en œuvre de procédures de secours onéreuses mais ne mettant pas en péril la vie de l'organisation.
- Vitales. Ce sont les données dont la perte peut être irrémédiable ou le coût de reconstitution, ou les risques juridiques et commerciaux auxquels ils exposent seraient prohibitifs pour l'organisation et susceptible de mettre en péril sa survie.

Identifier les données sensibles et préciser le niveau de protection à appliquer :

- Simple identification des accès,
- Accès réservé par sur-authentification,
- Chiffrement des données en stockage et dans les échanges.
- Hébergement répondant à des normes et homologations (données de santé par exemple).

Etablir les règles de production et de collecte.

Avec plus ou moins de détail et d'attention en fonction du degré de criticité, identifier les règles de qualité applicables à la production des données. Une donnée de mauvaise qualité ou incomplète équivaut à une donnée absente. Identifier également les points et les formes de collecte qui en garantisse la meilleure qualité.

Définir les modalités de sauvegarde, d'archivage et de destruction. Identifier les acteurs et les activités responsables de ces actions.

Préciser les usages légitimes des données vivantes, des sauvegardes et des archives et en déduire les droits d'accès correspondants.

Identifier les stades de conservation. Définir les modalités et les responsabilités pour les opérations de changement de stade.

Identifier les règles spécifiques aux échanges.

Mettre en place une infrastructure de conservation.

Les solutions de stockage, sauvegarde et archivage des données sont généralement des solutions mutualisées. Elles sont regroupées dans des infrastructures de conservation des données qui facilitent la transition d'un état à l'autre ou d'une solution à l'autre. Identifier tous les composants qui doivent être redondés et dupliqués y compris sur des sites distincts afin de

garantir le niveau de résilience aux pannes et atteintes approprié aux différents niveaux de criticité.

Etablir les responsabilités de mise en place et de maintenance.

Etablir les procédures de sureté et de reconstitution en cas d'altération ou d'indisponibilité.

Tester, étalonner et ajuster les délais de reprise d'activité.

Définir les modalités d'accès et de restauration.

Etablir les processus et les méthodes de restauration et de maintenance pour chaque point de traitement et de collecte.

Sécuriser les accès

Tester régulièrement le bon fonctionnement de ces procédures.

Auditer les processus.

A une périodicité régulière, vérifier le bon fonctionnement et la bonne application des procédures de collecte, de sauvegarde, de restauration, d'accès et de purge. Vérifier si le patrimoine de données est bien conforme aux prévisions

- Qu'il ne manque pas de données,
- Qu'il n'y ait pas trop de données,
- Qu'il n'y ait pas de données altérées,
- Que des données n'aient pas été accédées de manière illégitime.

Vérifier que les processus sont bien connus des parties en charge de les mettre en œuvre. Si besoin, mettre en œuvre les formations appropriées.

Mettre en œuvre les améliorations nécessaires.

Auditer l'infrastructure de conservation.

A une périodicité régulière, vérifier le bon fonctionnement et le bon niveau de maintenance en conditions opérationnelles de l'infrastructure de conservation des données.

Anticiper tout particulièrement l'obsolescence technique et la saturation de l'infrastructure. Vérifier les dispositifs de sécurité, de redondance et de réplication.

Envisager, si nécessaire des actions de changement de format ou de support et les mettre en œuvre.

Purger les données devenues inutiles.

Selon les procédures définies, supprimer les données devenues inutiles. Vérifier que cette suppression est bien totale et effective.

- Qu'il ne subsiste pas des copies quelque part,
- Pour certaines données sensibles, que l'effacement ne soit pas uniquement logique mais physique afin d'empêcher leur reconstitution.

Porter une attention particulière à la gestion des données en fin de contrat avec les sous-traitants.

Maintenir l'infrastructure et les données.

Assurer le niveau de ressource et de formation approprié pour le niveau de services attendu.

Mettre en place le pilotage et les mesures utiles à la prévention des risques de saturation, altération, voire pertes de données.

Mettre en œuvre les actions préventives et correctives et assurer le maintien en conditions opérationnelles.

Pour approfondir.