arbido

2017/3 Métadonnées – Données de qualité

Orban de Xivry Dominique, CEO de Rever S.A.

Adamski Muriel, chargée de communication de Rever S.A.

Garantir la bonne santé des données par la gestion des métadonnées

«L'excellence de l'information mène l'entreprise au succès». Ce slogan utilisé par nombre de vendeurs du monde informatique cache en réalité la nécessité d'une maîtrise permanente des données, de leur qualité et des règles de gestion qui les régissent.

Aujourd'hui la complexité des systèmes d'information et la volumétrie des données à gérer n'autorisent plus la simple «bonne volonté»: les défis imposés par les données commandent leur gestion rigoureuse et précise. L'expérience démontre que cet objectif ne peut être atteint que par l'utilisation «massive» des métadonnées au moyen d'automates.

Mais tout d'abord, un peu d'explication au sujet des métadonnées. Les métadonnées sont un ensemble d'informations qui décrit les données stockées. Ces métadonnées sont aux systèmes d'information ce que sont les cartes vis-à-vis des voies de circulation: une simple description. En fonction des objectifs poursuivis, les cartes décrivent soit des routes, soit des chemins pédestres, soit des voies navigables, soit le relief, soit encore des sites touristiques, ou toute combinaison de ces différentes alternatives.

> **«Les métadonnées sont aux systèmes d'information ce que sont les cartes vis-à-vis des voies de circulation.»**

Comme pour les cartes, les informations fournies par les métadonnées peuvent être très diverses: localisation des données, type de la donnée (document, image, email, fichier audio ou vidéo), règles de gestion, etc. Quels que soient le support et le format, les métadonnées offrent la possibilité de retrouver et de traiter de manière spécifique les données au moment voulu.

Indissociable du métier d’archivage, les métadonnées sont les héritières des systèmes de codification déjà pratiqués avant l’ère de l’informatique. Les bibliothèques, l’industrie de la musique, le monde de l’édition, les administrations sont de grands utilisateurs de métadonnées. Depuis l’arrivée d’Internet, l’utilisation des métadonnées s’est généralisée au sein des pages web (balises méta), des applications en ligne (réseaux sociaux, e-commerce, streaming) et leur introduction dans les processus de digitalisation permet à chaque utilisateur de maîtriser ses documents numériques.

Les métadonnées en entreprise

Pour les organisations, une trentaine d’années d’informatisation généralisée ont donné naissance à un parc de systèmes d’information et de méthodes de stockage des données extrêmement variés. Le cauchemar des silos de données éparpillés et opaques est une réalité et l’on ne compte plus les entreprises qui ont encore recours aujourd’hui à des systèmes archaïques qui ne permettent pas aux divers départements concernés de tirer pleinement profit de leur patrimoine informationnel, voire mettent en danger l’efficacité opérationnelle de l’entreprise.

> **«La déferlante numérique actuelle ne fait qu’accroître l’impérieuse nécessité d’une maîtrise des données.»**

La déferlante numérique actuelle ne fait qu’accroître l’impérieuse nécessité d’une maîtrise des données. Comme déjà dit, relever ce défi n’est possible que par la maîtrise des métadonnées qui génère une meilleure connaissance des systèmes d’information et décuple l’efficacité de l’utilisation des données.

Certains déplorent qu’un plan d’action pour la mise en place d’une gestion des données au travers des métadonnées est à la fois un travail fastidieux et un investissement inutile. L’expérience montre que le résultat est tout bénéfice: rapidité pour retrouver et traiter les données, élimination des redondances, découverte de liens cachés entre les données… ce qui permet de valoriser le «big data» disponible tout en utilisant le savoir-faire du personnel concerné dans l’entreprise.

Qualité des données et «golden record»

Une des utilisations des métadonnées les plus immédiates et les plus connues concerne la «qualité» des données. En effet, la connaissance de la localisation des données permet d'extraire et de comparer pour une même donnée les différentes versions stockées dans les multiples bases de données de l'entreprise. Ces comparaisons permettent de créer ce qu'il est convenu d'appeler le «golden record» ou «version unique de la vérité» qui permet de comparer et d'aligner les différentes versions. Pour illustrer ce propos par un exemple concret, prenons le cas de Madame X. Dans son compte client, il apparaît que celle-ci a 2 enfants, mais dans un autre système, elle a 3 enfants. Cette différence est facilement détectée via l'utilisation des métadonnées et permet au minimum d'uniformiser à 2 ou 3 le nombre d'enfants de Madame X. Est-ce pour autant le bon nombre? Seule une information complémentaire provenant probablement de Madame X elle-même permettra de s'assurer de la conformité des données à la réalité: elle a en fait 4 enfants.

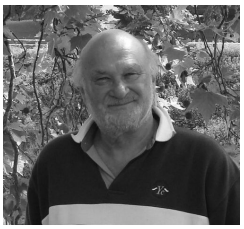
> **«Les métadonnées sont un élément stratégique pour la mise en place d'une réelle «gouvernance» des données.»**

Cet exemple illustre un point majeur du concept de «qualité» des données: à savoir la «distance» entre la réalité décrite par les données et la réalité du monde. D'un point de vue théorique, les données décrivent la réalité du monde avec un temps de retard et donnent une vue partielle de celle-ci. Les systèmes d'information sont donc par nature en retard et incomplets par rapport à la réalité qu'ils sont censés décrire, réalité qui est en perpétuelle évolution. Cette distance entre la description faite par les données et la réalité présente un «risque» potentiel pour l'entreprise qui prend des décisions en s'appuyant sur la réalité décrite. Dans ce contexte et en termes de gestion, il est primordial que la gestion de la «qualité» des données soit faite en fonction du rôle et de l'importance des données par rapport à des objectifs métiers.

Un métier d'avenir

On l'a vu, les métadonnées sont un élément stratégique pour la mise en place d'une réelle «gouvernance» des données. Elles facilitent l'intégration, le partage et la gestion quotidienne des données vitales pour l'entreprise, ce qui est la base du développement des stratégies «business». Pas de données exploitables sans métadonnées.

Même si le principe n'est pas nouveau, peu d'organisations ont décidé de privilégier les métadonnées par le passé (par manque de connaissance, par crainte des coûts ou par crainte de la durée de la mise en place). Toutefois, force est de constater que ces dernières années, le phénomène s'est inversé et que les «data engineers» ont vu leur rôle se renforcer considérablement pour devenir les moteurs de l'agilité et de la performance de l'entreprise dans un contexte de concurrence généralisée.



Dominique Orban de Xivry

Dominique Orban de Xivry a fondé la société belge Rever en association avec le laboratoire d'ingénierie des bases de données de l'Université de Namur. Il a également participé à la création de la société parisienne GEOLSemantics dont l'expertise tourne autour du profilage sémantique dans le cadre de l'analyse automatisée de textes.

Muriel Adamski

Muriel Adamski est chargée de la communication pour Rever depuis 2015. Après plusieurs années passées dans l'eCommerce, elle s'active maintenant à faire connaître la technologie de Rever qui est spécialisée en gestion des données et en modernisation des systèmes informatiques, un challenge passionnant au quotidien.

Résumé

Français
Les entreprises stockent de très nombreuses données qu'elles récoltent au fil des ans. Les métadonnées en simplifient la gestion, assurent le maintien de leur cohérence, permettent des analyses pertinentes et garantissent l'agilité de l'entreprise. En d'autres termes, les métadonnées permettent aux entreprises de passer du C.I.O. (Chief Information Officer) au C.I.O. (Chief Innovation Officer), c'est-à-dire le passage de la gestion des «containers» à la gestion des «contenus».

Deutsch
Unternehmen bewahren zahlreiche Daten auf, die sie im Lauf der Jahre gesammelt haben. Metadaten erleichtern dabei die Bewirtschaftung, stellen die Erhaltung der Daten und deren Konstistenz sicher, erlauben dem Unternehmen, flexibel reagieren zu können. Anders ausgedruckt: Die Metadaten erlauben es den Unternehmen, die Rolle des CIO vom "Chief Information Officer" zum "Chief Innovation Officer" weiterzuentwickeln und den Übergang von der Verwaltung von Behältern ("containers") zu Inhalten ("contenus") zu ermöglichen.