



La gestion de la qualité des données

La gestion de la qualité des données (ou Data Quality Management, DQM) devient une préoccupation croissante des directions des systèmes d'information. Initialement issue d'un besoin de mise en conformité des résultats des banques et des organismes financiers, la gestion de la qualité des données s'étend peu à peu à toutes les fonctions opérationnelles des entreprises des différents secteurs économiques. En effet, ces dernières ont pris conscience que leur efficacité et leur compétitivité dépendent largement de la valeur et de la qualité des informations qu'elles gèrent. Toutefois, un investissement massif dans des solutions logicielles ne suffit plus. Désormais, il leur faut s'intéresser également au contenu. Selon la société d'études spécialisée The Data Warehousing Institute, la mauvaise qualité des données coûterait aux entreprises américaines plus de 600 milliards de dollars chaque année.

Qu'est-ce que la qualité des données ?

Dans l'industrie, la qualité d'un produit est correctement appréciée en comparant ses effets aux attentes des clients. Un produit ou service répondant parfaitement aux besoins des clients est donc réputé de grande qualité. Il en est de même pour les données : une donnée est de qualité si elle répond parfaitement aux besoins des utilisateurs de cette donnée.

Si la justesse est une des conditions évidemment nécessaires à la définition d'une donnée de qualité, d'autres caractéristiques doivent également être prises en compte. Comme l'indique le tableau suivant, les données peuvent être appréciées en fonction de leur contenu, leur accessibilité, leur flexibilité, et leur sécurité.

PRINCIPALES CARACTERISTIQUES DE QUALITE DES DONNEES	
Qualité du contenu	<ul style="list-style-type: none"> • Justesse • Adéquation aux besoins • Compréhension
Accessibilité	<ul style="list-style-type: none"> • Disponibilité • Facilité d'accès
Flexibilité	<ul style="list-style-type: none"> • Évolutivité • Cohérence avec d'autres sources • Traduction
Sécurité	<ul style="list-style-type: none"> • Respect de la confidentialité • Fiabilité • Traçabilité • Intégrité

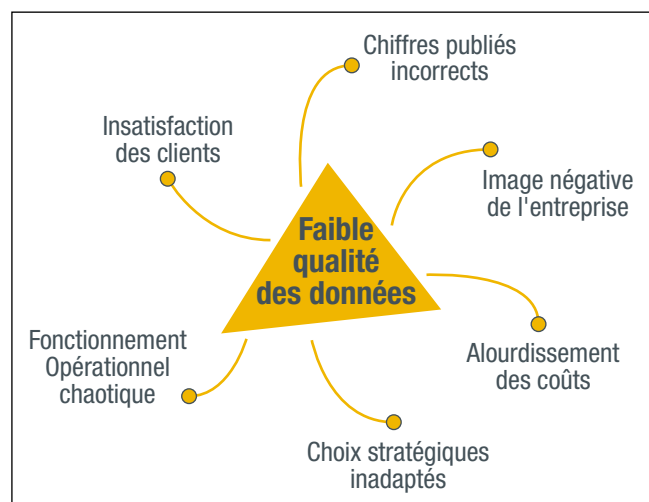
Principales caractéristiques de qualité des données

En termes de couverture fonctionnelle, les applications métier du système opérationnel, tout comme celles d'aide à la décision du système de pilotage sont concernées. Il est à ce propos intéressant de souligner que si les données du système opérationnel sont de mauvaise qualité, celles du système de pilotage ont toutes les chances de l'être aussi, puisque les données de celui-ci sont généralement issues du système opérationnel. Résultat : opérationnel hasardeux et pilotage de l'entreprise erroné.

Notons aussi que l'urbanisation du système d'information et les architectures de type SOA nécessitent plus que jamais de s'intéresser aux données et à leur qualité, avec la mise en place ou l'alignement de référentiels de données.

Ainsi, l'adresse postale demeure une des données les plus utilisées par les entreprises, notamment pour leur gestion de leurs partenaires clients et fournisseurs. C'est aussi l'une des plus touchées par les défauts de qualité. Des efforts particuliers méritent donc d'être apportés pour améliorer la gestion de cette donnée complexe.

Le chaos ruineux de la non-qualité des données



Principales conséquences de la non-qualité des données

Les conséquences de la faible qualité des données sont nombreuses et reviennent cher aux entreprises. Pire encore, l'impact se propage jusqu'à l'extérieur, notamment en raison de la place grandissante des échanges électroniques. Au final, la détérioration de l'image et la perte de crédibilité s'avèrent parfois fatales aux organisations dont les données sont de mauvaise qualité. Les conséquences rencontrées le plus fréquemment sont les suivantes :

- **Insatisfaction des clients** : les clients victimes de données incorrectes pardonnent difficilement. Une livraison tardive suite à une erreur sur la date de livraison reste une erreur courante. Avec Internet, la sentence est encore plus rapide. D'ailleurs, un internaute s'apercevant que les données d'un site sont peu fiables ou incorrectes -comme des prix non actualisés- n'y reviendra jamais.
- **Non-conformité des chiffres publiés** : des erreurs dans les données comptables entraînent généralement de lourdes sanctions (amendes et peines). Et ce, même si cette situation ne résulte pas d'une manipulation volontaire. Suite aux récentes affaires de falsification des comptes, toute erreur dans la publication des chiffres paraît désormais suspecte.
- **Dévalorisation de l'image de l'entreprise** : en conséquence du point précédent, des erreurs portant sur des données diffusées à l'extérieur de l'entreprise détériorent rapidement l'image de celle-ci, et handicapent son expansion commerciale. De plus, l'impact boursier des sociétés cotées est immédiat, entraînant la chute rapide du cours de l'action.
- **Perturbation du fonctionnement opérationnel** : les données incorrectes au sein de l'entreprise ralentissent le fonctionnement opérationnel, et rendent les décisions difficiles ou incertaines. Cela génère aussi du stress et de la démotivation chez les collaborateurs. De plus, le surcroît de travail engendré par des erreurs de données n'est jamais le bienvenu. Enfin, la perte de confiance dans les informations manipulées n'est propice ni à l'initiative, ni à un climat sain.
- **Erreurs de stratégie** : définir une stratégie nécessite plus de temps si les informations utilisées sont de mauvaise qualité ou tout simplement erronées. Comment réfléchir à l'avenir

si la situation présente n'est pas maîtrisée ? Les dirigeants peuvent évidemment se fier à leur intuition, mais ce type de démarche montre rapidement ses limites.

- **Augmentation des coûts** : les impacts énoncés précédemment engendrent évidemment des surcoûts. Les situations variant d'une entreprise à une autre, il est très difficile d'estimer le coût de la non-qualité. Certains experts estiment cependant que ce coût caché peut représenter plus de 10 % du chiffre d'affaires, ce qui ampute de façon conséquente la rentabilité.

Les origines du mal : enjeu sous-évalué et mauvais traitements

Les causes d'anomalies portant sur les données sont multiples et variées, mais elles s'expliquent le plus souvent par une sous-estimation générale de ces enjeux. Les décideurs négligent souvent les tâches d'initialisation des données pourtant capitales, tant au démarrage d'une nouvelle application (reprise des données en masse), qu'en régime permanent (saisie manuelle au fil de l'eau). Enfin, certains défauts de qualité trouvent aussi leur origine dans la conception des applications informatiques, et dans les défaillances logicielles (bugs).

La **sous-estimation de l'enjeu des données** tient au fait que l'objectif d'une entreprise ne semble pas lié à la production des données. Bien que l'exercice des différents métiers nécessite des informations, le but premier consiste plutôt à obtenir, selon le cas, un client satisfait, un niveau de vente supérieur, un meilleur produit, une meilleure marge, etc. Les données ne se placent donc pas au premier plan de l'activité, et ne sont généralement pas perçues comme un élément essentiel de compétitivité. D'ailleurs, un grand nombre de managers considère que l'investissement dans une politique qualité des données se traduit uniquement en perte de temps et d'argent.

La **reprise des données** d'une ancienne application vers une nouvelle génère souvent l'absence de qualité, car les décideurs ne lui accordent pas toujours l'importance qu'elle mérite. Pourtant, la reprise des données en masse peut véritablement être assimilée à un transfert de patrimoine. En effet, les données que l'on initialise dans le nouveau système proviennent généralement d'un ancien système. Or si les données de l'application d'origine affichent une qualité médiocre, un transfert effectué sans opération de nettoyage aboutira à une application cible avec les mêmes défauts de qualité. Cela est particulièrement vrai pour les données stables, comme la base Clients, Fournisseurs ou encore Articles. Ainsi, il est hélas fréquent de transférer des doublons ou des données obsolètes d'une application vers une autre. En conséquence, la nouvelle application ne produit évidemment pas tout le bénéfice que l'on pourrait en attendre.

Au quotidien, la **saisie des données** au fil de l'eau présente aussi des lacunes. Malgré des contrôles automatisés mis en place dans les applications, la méconnaissance des règles de gestion et le manque de formation engendrent souvent des saisies incorrectes. De nombreux utilisateurs créent et gèrent des données utiles à leur travail, mais peu se préoccupent de l'utilisation qui en est faite en aval. Ce cloisonnement nuit considérablement à la qualité et à la fluidité de l'information.

Les **défauts de conception** et les **anomalies des logiciels**

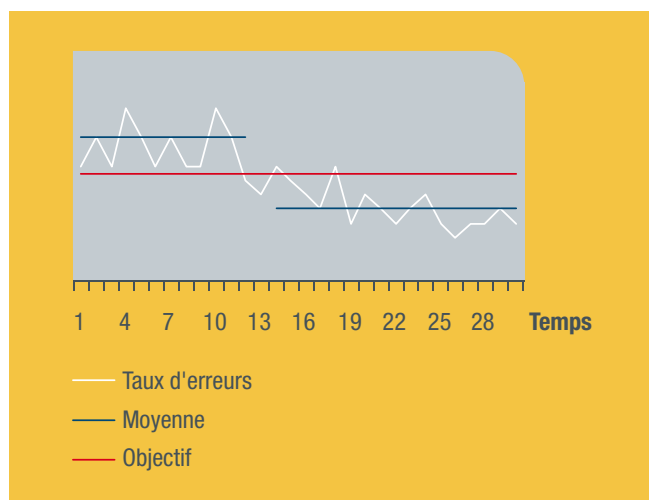
(bugs) portent aussi leur part de responsabilité. Bien que le coût correspondant soit loin d'être négligeable, surtout s'il s'agit d'un problème de conception, la gravité vis-à-vis des données est à relativiser. Car une fois corrigée, l'application concernée n'est plus directement à l'origine des problèmes de qualité de données.

Un problème directement lié aux métiers de l'entreprise

La prise de conscience des dirigeants de la relation entre qualité des données et performance de l'entreprise, est très certainement la première source de progrès à développer. Il est intéressant de constater que la responsabilité de la qualité des données est actuellement, dans le meilleur des cas, entre les mains de la direction des systèmes d'information. Or celle-ci ne peut évidemment pas être tenue responsable de tous les problèmes de données, tout simplement parce qu'elle ne contrôle pas tout, et n'est généralement pas à la source de la création des données. La qualité des données n'est pas une problématique technique, elle touche avant tout au métier de l'entreprise.

Partant de là, la mise en place de **programmes d'amélioration de la qualité de l'information et des données** est l'un des vecteurs clés de progrès. Un programme doit être géré comme un projet et impliquer largement la ou les directions fonctionnelles concernées (vente, achats, production...). Bien entendu, un tel projet ne doit être lancé que s'il y a un réel besoin et un enjeu significatif pour l'entreprise. Il est donc recommandé de bien s'assurer que les processus concernés jouent un rôle significatif dans la compétitivité de l'entreprise. Il convient également de définir un objectif clair et réaliste et d'évaluer les chances d'aboutir à un résultat positif. La solution mise en œuvre dans le cadre d'un programme peut comporter des aspects techniques et/ou organisationnels.

Autre point important, l'entreprise doit se doter de moyens de mesure pour évaluer la qualité des données avant et après mise en œuvre du programme. Par exemple, la mesure du taux d'erreurs au sein d'un processus métier peut être l'un de ces indicateurs.



Mesure du taux d'erreurs avant et après mise en œuvre de la solution



Outre le lancement de programmes d'amélioration ponctuels, il s'avère très profitable d'instaurer de façon durable de bonnes pratiques de data management (gestion des données). Basé sur des principes clairs privilégiant les mesures préventives et le pragmatisme, le data management permet incontestablement d'améliorer la qualité des données et de rendre l'entreprise plus performante et compétitive. Les bonnes pratiques recouvrent l'organisation, la gestion, la codification, la documentation, la surveillance, et l'auditabilité des données.

DQM et MDM : les outils à qualités

Pour supporter la gestion de la qualité des données, **des outils de DQM** (Data Quality Management) sont désormais disponibles. Selon le Gartner Group, ce marché représente aujourd'hui dans le monde 300 millions de dollars, et devrait croître de 18 % chaque année pour atteindre 677 millions de dollars en 2011. Le marché actuellement tiré par les pays anglo-saxons commence seulement à toucher la France. D'ailleurs, la plupart des fournisseurs sont américains, à l'instar de DataFlux, Trillium Software, IBM ou Informatica. Business Objects, racheté récemment par l'Allemand SAP, fait figure d'exception.

L'objectif de ces outils consiste à améliorer la qualité des données. Une majorité de solutions permet de détecter des anomalies, pour les corriger ensuite. Il s'agit donc dans l'ensemble d'un mode curatif, même si certains produits fonctionnent de manière préventive en détectant les anomalies à la source en temps réel. En pratique, ces outils s'avèrent très efficaces. Ils permettent en particulier d'améliorer la standardisation de certaines données comme les adresses, de détecter et d'éliminer les doublons, de produire des statistiques et des rapports facilitant la détection d'anomalies.

Outre ces outils très spécialisés, on assiste depuis quelques années à un engouement pour les **solutions de MDM** (Master Data Management). Le MDM permet de gérer au sein d'un référentiel un certain nombre d'objets métiers appelés données de base (principalement articles / produits, clients, fournisseurs, employés) pour les diffuser aux différentes applications de l'entreprise. Grâce au référentiel commun, les utilisateurs peuvent manipuler la même version de la réalité, ce qui facilite évidemment les échanges entre applications, et améliore la qualité des flux métiers sous-jacents.

Les solutions de MDM améliorent évidemment la qualité des données, grâce à la gestion mieux maîtrisée des informations

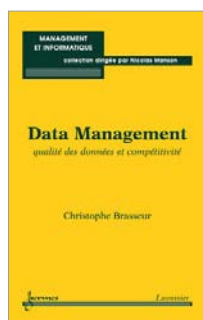
de base de l'entreprise. Parmi les fournisseurs de MDM, on retrouve les grands éditeurs d'ERP tels que SAP (avec la solution NetWeaver MDM) qui est un pionnier. IBM est aussi très bien positionné sur le marché avec les produits WPC (gestion des produits) et WCC (gestion des clients) intégrés à son offre WebSphere. Par ailleurs, les acteurs de la Business Intelligence développent des offres MDM, comme Hyperion ou Teradata. Enfin, une multitude de « pure players » tels que Sunopsis ou FullTilt Solutions apportent des solutions intéressantes. Sans oublier les éditeurs de bases de données comme Oracle (Enterprise Master Data Management et Hyperion MDM) ou Microsoft (rachat de Stratature en juin 2007).

Au-delà des outils, et c'est vraisemblablement l'un des points essentiels, la prise de conscience des enjeux, et l'adhésion des collaborateurs à un programme d'amélioration de la qualité est indispensable. La communication dans l'entreprise autour du thème de la qualité des données et la prise en compte de l'aspect humain constituent incontestablement les clés de la réussite de ce type d'approches.



Christophe Brasseur,
Consultant Manager
chez Caggemini

Christophe Brasseur est ingénieur ESTP et titulaire d'un DESS de gestion de l'IAE de Montpellier. Il a participé à de nombreux projets en France et à l'étranger en tant que directeur de projet ou consultant, principalement dans le secteur des services publics, du pétrole, de l'énergie, des utilités, et de la chimie. Il est actuellement Consultant Manager chez Caggemini.



Christophe Brasseur :
« Data Management, qualité des données et compétitivité »,
éditions Hermès-Lavoisier, Paris, 2005.