Des Données de Qualité

Exploitez le capital de votre organisation

Un livre blanc de JEMM research
Janvier 2008

Table des matières

Table des figures

Introduction

Dans un contexte où les défis des entreprises et administrations sont de plus en plus nombreux, disposer d'un capital de données de qualité devient une nécessité incontournable. Déferlement d'informations sans précédent, pressions réglementaires, exigences de contrôle interne, cohérence des échanges avec les partenaires, satisfaction des clients sont autant de défis à relever par les entreprises. La maîtrise de la qualité des données est désormais un enjeu important. Il s'agit de fournir des données correctes, complètes, à jour et cohérentes tout en mettant en place des indicateurs compréhensibles, faciles à communiquer, peu coûteux et simples à calculer. La direction générale et ses directions métiers doivent disposer d'une vision unifiée et exploitable des informations, afin de prendre les bonnes décisions au moment opportun.

Pourtant, la gestion de la qualité des données reste essentiellement du domaine de la direction informatique. Historiquement, les systèmes d'informations ont conçu des applications pour traiter les données élémentaires de l'entreprise. Aujourd'hui, les directions métiers demandent à l'informatique de leur mettre à disposition des tableaux et indicateurs basés sur ces traitements et ces applications. Mais que se passe-t-il si les données issues des applications sont erronées, tronquées ou redondantes… ? La direction informatique peut-elle connaître les règles métiers associés au domaine fonctionnel ? Une réponse technologique n'est pas suffisante et il est clair que les directions métiers doivent aussi s'impliquer fortement dans cette gestion.

Les responsables fonctionnels et les équipes de la direction informatique doivent mettre leur force en commun pour développer un programme de gestion de la qualité des données. Mais avant de se lancer dans l'aventure, plusieurs questions se posent. Comment réaliser des référentiels au service de la qualité des données clients, fournisseurs et partenaires ? Quelles sont les bonnes pratiques en matière de gestion des données ?

Ce livre blanc décrit la problématique de la qualité des données du point de vue des directions métiers. Au-delà de la simple vue informatique, il explicite l'impact métier d'un manque de données de qualité. Il décrit des moyens de mise en place d'une politique de qualité des données et présente des scénarios de mise en œuvre de cette politique.

Le capital de l'entreprise

Aujourd'hui, l'entreprise privée ou publique aujourd'hui est confrontée à un défi de flexibilité. D'un coté, elle doit s'adapter rapidement à un environnement changeant, dans lequel le temps et les distances diminuent, les opportunités doivent être saisies immédiatement, les réglementations s'accumulent. Dans ce contexte non stabilisé, l'entreprise est confrontée au défi de l'adaptation permanente : détecter les fenêtres d'opportunités afin de bénéficier d'un avantage concurrentiel, augmenter l'innovation pour différencier sa proposition de valeur par rapport à la concurrence, analyser son environnement afin de prendre les bonnes décisions au bon moment, rationaliser son organisation et ses processus internes pour améliorer sa productivité, intégrer les interactions avec ses partenaires et fournisseurs afin de réduire les délais et de faciliter les processus.

De l'autre, cette nécessité de rapidité et de flexibilité doit reposer sur un environnement rigoureux. En effet, l'entreprise doit s'appuyer sur une gouvernance, un ensemble de règles de prises de décision, de transparence et de surveillance qui lui permettent de contrôler son fonctionnement. Le comité exécutif doit pouvoir prendre les décisions stratégiques en s'appuyant sur des éléments fiables. Il est important de minimiser la prise de risque en mettant en place des moyens effectifs de contrôle. Le pilotage de l'organisation nécessite la manipulation d'indicateurs fidèles et à jour de l'activité. Enfin, l'entreprise doit pouvoir justifier et garantir sa conformité aux réglementations, lois et régulations.

Au cœur de ce système complexe, cette organisation aux multiples facettes, sont les données que l'entreprise récolte, génère, manipule, alimente et publie. Clients, produits, fournisseurs, transactions de vente : toutes ces entités doivent être décrites et suivies d'une manière ou d'une autre. C'est à partir de ces données que l'entreprise évalue les opportunités qui se présentent à elle. La direction générale base ses décisions sur leur analyse exhaustive. Tous les collaborateurs les manipulent dans le cadre de l'exécution de leurs tâches et activités. Les partenaires les synchronisent avec leurs données internes afin de coordonner les actions. Les clients les consultent pour interagir avec l'entreprise. Enfin, les observateurs les analysent pour évaluer la santé financière et économique de l'entreprise. Avec les technologies de l'information, la sophistication de l'usage des données par les entreprises et les gouvernements s'est développée de manière exponentielle. Les fournisseurs de solutions technologiques ont créé beaucoup de termes, de concepts, de produits pour répondre à ce besoin : système d'aide à la décision, entrepôt de données, gestion de la relation client, business intelligence, gestion des données de référence (MDM). Mais, en tout état de cause, le besoin reste fondamental, les organisations doivent optimiser l'usage de leur données pour augmenter leur efficacité d'aujourd'hui et améliorer la stratégie de demain.

Pourtant, la qualité des données est rarement une priorité pour les organisations. Il est fréquent d'entendre des remarques telles que « Nos données sont de qualité suffisante » ou « On les nettoiera plus tard ».

Le coût de la non-qualité

Les données contribuent au succès de l'activité de l'entreprise. Leur qualité représente donc un enjeu critique pour l'entreprise dans les trois étapes de leur cycle de vie :

- lors de la saisie
- au cours des transformations et agrégations
- pendant l'analyse et la présentation des résultats

Saisir des données de qualité

Les entreprises doivent aujourd'hui faire face à un déferlement de données et d'informations sans précédent. On estime que davantage d'informations ont été générées pendant ces 30 dernières années que lors des 500 précédentes. Dans le monde de la finance, le nombre de transactions boursières double tous les 6 à 9 mois. Dans le monde de la logistique, la capacité de stocker des informations granulaires grâce à la technologie RFID génère un volume entre 10 et 100 fois plus important que celui de la technologie de code-barres. Ces données à l'état brut alimentent les systèmes d'information. Mais la gestion de leur qualité n'est toujours pas soumise à des règles et des standards. Cela conduit à des prises de décision à partir de données incorrectes ou mal interprétées. Aux Etats-Unis, une étude[1] a estimé le coût de la mauvaise qualité des données à plus de 600 milliards de dollars pour les entreprises chaque année.

La mauvaise qualité des données est due principalement aux erreurs de saisie de l'information à la source. Fautes d'orthographes, codes incorrects, abréviations erronées, saisies dans un mauvais champ sont autant de sources de dégradation de la qualité qui peuvent avoir des conséquences néfastes pour l'entreprise.
Les silos applicatifs traditionnels, les fusions et acquisitions d'entreprises entrainent une duplication des données dans les systèmes d'informations. Lors de l'intégration des silos ou de la consolidation des applications, on retrouve des données enregistrées plusieurs fois dans les systèmes informatiques sous des identifiants différents. De plus, des données exactes à un moment donné, peuvent devenir erronées à la suite d'un changement de situation, le déménagement d'un client peut amener la création d'un nouvel identifiant au lieu d'une modification de la fiche initiale.

Tokyo Stock Exchange

Le deuxième groupe bancaire japonais, Mizuho, a perdu 286 millions d'euros …. pour une faute de frappe.
En décembre 2005, lors de l'introduction en bourse d'une petite société, J-Com, un courtier de cette banque avait placé 610 000 titres à 1 yen au lieu de vendre 1 titre à 610 000 yens. L'erreur n'avait pas pu être rattrapée à temps par les services informatiques de la bourse de Tokyo (TSE).
Le patron de TSE, Takuo Tsurushima, démissionna un mois plus tard.

De même, la saisie des données directement par le client peut avoir des conséquences pour le moins inattendues. Dans le domaine du tourisme, la figure 1 montre qu'il est possible de réserver sur Internet un vol aller/retour avec le vol retour décollant avant l'arrivée du vol aller ! Il faut certes être étourdi… Mais le contrôle de la validité de la proposition n'est-il pas de la responsabilité des compagnies aériennes ? D'ailleurs, on pourrait poser la question aux trois touristes norvégiens de la compagnie aérienne qui les a amenés à Rodez alors qu'ils désiraient passer des vacances sur l'île grecque de Rhodes[2] ?


1 Recherche 2 Résultats 3 Informations passagers & Paiement 4 Confirmation

Vols Low Cost - Réservation

Votre réservation

Détails du vol

Aller :	mercredi 5 décembre 2007	
Départ : 09h30 Paris Beauvais Arrivée : 11h30 Rome Ciampino		Vol Low Cost 9632 Classe Economique.
Retour :	mercredi 5 décembre 2007	
Départ : 11h20 Rome Arrivée : 13h30 Paris		Vueling 9981 Classe Economique.

Récapitulatif de votre commande

Descriptif	Quantité	Prix par passager	Frais de dossier par passager	Prix total
Tarif Adulte	1	124.03 €	18.00 €	142.03 €
Montant total				142.03 € TTC

Figure 1 - Saisir des données de qualité

[1] *"Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data" - The Data Warehousing Institute - 2002*
[2] afp.google.com/article/ALeqM5iuffqw1CR1lBQI5qOYmc6J-xmc2w

Exploiter des données de qualité

Dans sa démarche de flexibilité, l'entreprise recherche l'efficacité opérationnelle. L'exploitation de données de qualité permet d'optimiser la participation et les interactions entre tous les collaborateurs au-delà des frontières administratives ou techniques. Pourtant, beaucoup d'entreprises négligent d'analyser la qualité de leurs données, ce qui les conduit à exploiter des données fausses ou erronées. Les silos applicatifs restent nombreux, rendant difficile le partage et l'intégration des données. Cela entraine de nombreux impacts sur le pilotage et la performance de l'entreprise. Les exemples sont nombreux dans tous les secteurs d'activité aussi bien au sein des entreprises privées que publiques.

En 1999, la NASA[3] a perdu le satellite Mars Climate Orbiter à cause de données erronées. En effet, le satellite fut détruit pendant sa mise en orbite autour de Mars à une altitude de 50 km de la surface (l'altitude normalement prévue était de 150 km) par les turbulences et les frottements atmosphériques. L'enquête a mis en évidence que certains paramètres avaient été calculés en unités de mesure anglo-saxonnes et transmises telles quelles à l'équipe de navigation, qui attendait ces données en unités du système métrique. Cette « petite » erreur a coûté 125 millions de dollars aux contribuables américains.

Plus près de nous, les difficultés de l'Airbus A380 concernant la phase d'industrialisation de l'avion ont porté essentiellement sur le câblage électrique d'une partie du fuselage, la conception ayant été menée avec des logiciels de versions différentes pour la partie française et la partie allemande[4].

Il est toujours possible de chiffrer le coût direct de la non-qualité des données. La Figure 2 fait le calcul du retour sur investissement d'une campagne marketing d'un opérateur téléphonique.

Hypothèses		
Nombres de brochures envoyées	50000	
Coût total du programme	150 000 €	
Bénéfice moyen par vente	1 000 €	
Ratio de duplication	**1**	**1.04**
Taux de réponse	2 %	1.92 %
Ratio de foyer	**1**	**1.11**
Taux de conversion	20 %	18.02 %
Résultats		
Nombre de réponse	1000	962
Cout par réponse	150 €	156 €
Nombre d'acheteurs	200	173
Bénéfice total de la campagne	200 000 €	173 250 €
Retour sur Investissement	**33.33 %**	**15.50 %**

Figure 2 - Valeur de la qualité dans une campagne marketing

[3] www.cnn.com/TECH/space/9909/30/mars.metric/

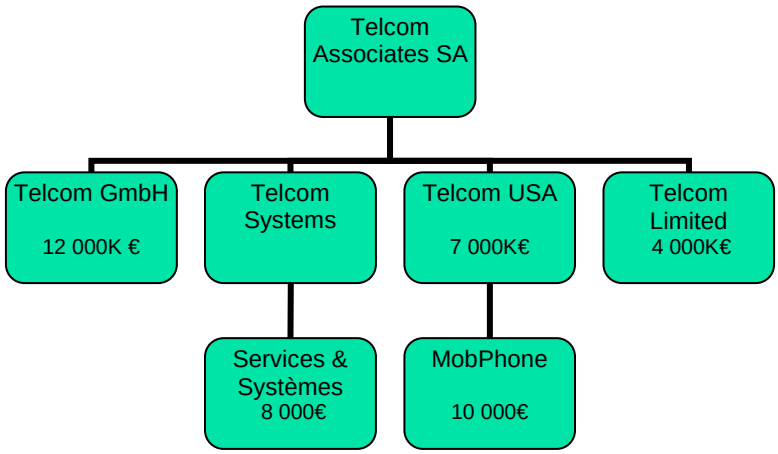
[4] Audition du 22 novembre 2006 de M. Louis Gallois, co-Président exécutif d'EADS devant la commission des affaires étrangères, de la défense et des forces armées du sénat.

L'entreprise veut faire l'envoi d'une brochure annonçant un nouveau service à l'ensemble de ses clients. Sa base Client contient des enregistrements dupliqués (ratio de duplication 1.04) ainsi que des enregistrements multiples pour le même foyer (ratio de foyer 1.11). Ce simple calcul montre qu'une opération marketing sur des données de qualité peut doubler son efficacité.

Analyser des données de qualité

Enfin, dans l'analyse et la présentation des données, l'entreprise doit en garantir la qualité. L'impact est important. Le meilleur des tableaux de bord, l'analyse décisionnelle à l'aide d'outils de business intelligence les plus sophistiqués, ne peuvent donner des éléments de prises de décision fiables que si les données ayant servi à générer le tableau ou l'analyse sont correctes, cohérentes et à jour. Premier employeur en France, l'administration française a créé un observatoire de l'emploi public pour « assurer la cohérence des sources statistiques ». Dans son rapport de 2005, l'observatoire dénombrait en France entre 5,14 millions (dans une approche juridique) et 5,88 millions (dans une approche économique) d'agents dans la fonction publique[5]. On peut également citer le cas d'une grande compagnie d'assurance qui avait décidé de fusionner ses bases de données Clients pour avoir une meilleure compréhension des ses clients et des produits qu'ils achetaient, afin d'améliorer son offre de services. Avant le projet, le management pensait avoir 13 millions de clients, estimation basée sur les informations disponibles. Lors du projet, les équipes ont découvert beaucoup d'enregistrements dupliqués dans les bases et il a fallu réduire de 5 millions le nombre de clients de la compagnie à la fin du projet.

Un processus d'analyse de risque peut être rendu inopérant à cause de données non fiables. Il est important de connaître les liens juridiques qui existent entre les sociétés de votre base Clients. Un risque limité sur l'encours de la société Services & Systèmes peut devenir très important une fois consolidé au niveau de la structure Telcom Associates SA. De même, lancer une procédure juridique pour un encours de 10 000 € sur la société MobPhone peut avoir des impacts critiques dans les relations stratégiques avec la société mère qui génère un chiffre d'affaires de plus de 1 million d'euros.



La récente crise des subprimes a provoqué une crise de confiance générale dans le système financier, une chute des marchés financiers et une crise de liquidité bancaire. Cette crise, ajoutée aux scandales financiers précédents et aux faillites d'entreprises comme Enron en 2001, a justifié le besoin de mettre en œuvre des procédures de respect et de conformité aux régulations. Ces lois et régulations ont renforcé la responsabilité juridique et pénale des dirigeants. La direction financière doit aujourd'hui produire en temps voulu, des rapports reflétant la situation financière exacte de l'entreprise. Il faut donc mettre en place des processus de contrôles internes tout en réduisant les coûts additionnels et la complexité de création de ces rapports. En outre, les procédures d'audit requièrent de pouvoir justifier l'état des données utilisées pour produire ces rapports réglementaires. La qualité des données utilisées prend ici aussi toute sa valeur.

[5] Analyse de l'emploi public et de son évolution – Observatoire de l'emploi public - 2005

Un enjeu stratégique

« Aujourd'hui, 16 % des entreprises ont mis en place un programme de qualité de données »

Depuis de nombreuses années, les DSI reconnaissent l'importance de la qualité des données comme élément fondamental de leur stratégie de gestion de l'information. Toutefois, il leur est difficile de mettre en place les procédures et les programmes adéquats. Une étude récente[6] montre qu'à ce jour seulement 16 % des entreprises ont mis en place un programme de qualité de données. En revanche, dans une projection à trois ans, près de 80 % en auront un.

Dans le chapitre précédent, nous avons vu que la non-qualité des données avait des impacts très importants sur la performance de l'entreprise. Ainsi interrogées[7], les entreprises dans leur grande majorité identifient des problèmes de confiance limitée des utilisateurs, de mauvaise productivité, de prises de décision plus difficiles, et de coût de possession plus élevé, comme conséquences directes de la non-qualité de leurs données. Il est clair que la mise en place d'un programme d'amélioration de la qualité des données apportera des bénéfices importants pour les organisations. Il faut maintenant convaincre la direction générale et les directions métiers de l'utilité d'un tel programme. En effet, la qualité des données est rarement une priorité pour les directions métiers. Il est fréquent d'entendre des remarques telles que « Nos données sont de qualité suffisante » ou « On les nettoiera plus tard ».

Cette différence d'appréciation de l'importance de la qualité des données vient peut-être de la différence de point de vue entre les directions métiers et l'informatique. Chacun a des priorités différentes et personne n'est responsable de la problématique globale.

Un enjeu de l'informatique seulement ?

L'entreprise a besoin de son système d'information pour supporter sa stratégie, ses processus et son développement. Parties intégrantes de l'organisation, les applications et systèmes devenus très complexes gèrent des volumes de données considérables difficiles à manipuler. Les données gérées sont dynamiques et changent souvent. L'intégration des sources de données extérieures émanant des partenaires n'a pas simplifié les opérations. Les nouveaux canaux Internet d'interaction avec les clients ont multiplié les risques potentiels de corruption. Le respect des lois, en particulier celles sur la protection des données personnelles, exige de mettre en place des mécanismes stricts de contrôle de l'intégrité des données. Il en résulte que, pour de nombreuses directions métiers, l'origine et la fiabilité des données ne sont plus toujours clairs. Du côté des métiers, le problème est essentiellement identifié comme « informatique ». On entend souvent des remarques telles que « Ces données viennent de l'application, elles doivent être correctes », « Les données sont erronées. Je vais demander au département informatique de les corriger » ou « Voilà une technologie qui va me garantir des données de qualité ».

En revanche, la direction informatique n'est pas directement impactée par la mauvaise qualité des données. Les programmes et les procédures ne peuvent que gérer les données sans pouvoir garantir leur qualité. C'est le principe « *Garbage In- Garbage Out* », anglicisme pouvant se traduire par "déchet à l'entrée, déchet à la sortie", utilisé pour qualifier le fait que dans le domaine des données, de mauvaises données au départ de la chaîne ne peuvent générer que de mauvais résultats en fin de course. En effet, la DSI ne possède pas la connaissance et l'expertise des activités gérées par les directions métiers. Elle ne connaît pas les règles métiers associées aux données. La réponse technologique seule n'est en aucun cas suffisante pour garantir la qualité des données.

[6] Accenture CIO Survey 2007
[7] 2006-07 Scorecards for Data Governance in the Global 5000 – CDI Institute

L'initiative Qualité des Données

Pour exploiter au mieux son capital de données, l'entreprise doit lancer une initiative Qualité des Données. Stratégique pour l'entreprise, c'est un programme permanent et pas une mesure isolée dans le temps. Elle concerne de nombreuses fonctions métiers et informatiques dans l'entreprise. Elle nécessite de définir des processus formels de qualité des données appliqués par la direction informatique et les directions métiers. Cette initiative est supportée par des solutions technologiques qui permettent d'agir sur l'ensemble des projets : de la standardisation des données, au profilage, en passant par le nettoyage, jusqu'à l'enrichissement.

Il faut définir les règles de gestion des données de l'entreprise. Elles sont édictées pour garantir la qualité de complétude, conformité, cohérence, exactitude, non-duplication et intégrité des données. C'est le programme de gouvernance des données. Pour garantir son efficacité, ce programme doit inclure un comité, un ensemble de procédures et leur plan d'exécution. Autour des données, le programme doit mettre en pratique les contrôles de qualité de données et diffuser les bonnes pratiques. Il est articulé autour de deux approches :

- Une approche proactive incluant les bonnes pratiques à appliquer quand de nouvelles données sont générées, de nouveaux projets affectant les données sont lancés ou des actions de qualité sont effectuées
- Une approche réactive incluant les actions de correction de non qualité, les actions de mise en conformité suite à de nouveaux règlements, les actions d'intégration suite à la mise en œuvre des nouvelles architectures orientées-services (SOA).

D'une manière générale, l'initiative Qualité des Données doit couvrir les aspects suivants :

- Définition les objectifs de qualité des données
- Définition, mise en place et gestion des standards de qualité
- Vérification de la conformité réglementaire avec les standards de qualité qui ont été définis
- Identification des domaines d'amélioration de la qualité des données
- Mise en place des indicateurs de qualité des données
- Mesures et publication des rapports pour le management
- Sensibilisation et formation des équipes sur la problématique de qualité des données

La méthode

Il existe de nombreuses méthodes pour identifier, mesurer et résoudre les problèmes de qualité des données. Souvent, les entreprises ont développé de manière empirique des programmes d'amélioration de la qualité des données pour répondre à un problème critique à un moment donné. Les praticiens et les universitaires se sont penchés sur le problème de la qualité en général et des données en particulier et ont développé des méthodologies adéquates. On peut citer ici la méthodologie TIQM[8] d'InfoImpact ou le programme TDQM[9] développé et enseigné au Massachusetts Institute of Technology.

Toutes les méthodes d'amélioration de la qualité de données comprennent un cycle de quatre étapes :

- Définition
  Dans cette étape, l'entreprise définit comment mesurer la qualité des données afin de répondre aux besoins des utilisateurs. Elle décide des axes prioritaires de travail.

[8] www.infoimpact.com/tiqmmethodology.cfm
[9] web.mit.edu/tdqm/

- Mesure
  Il faut maintenant mesurer la qualité des données dans les projets en ligne avec la stratégie de l'entreprise et suivant des critères et des mesures définis par les utilisateurs.
- Analyse
  L'organisation évalue l'impact et les coûts de la non-qualité pour les directions métiers. Elle prépare aussi les plans d'amélioration de cette qualité. L'objectif est de présenter aux responsables concernés le business case du projet d'amélioration.
- Amélioration
  Dans cette étape, l'entreprise exécute les projets d'amélioration et de correction. Elle met en place les outils de mesure. Elle vérifie les indicateurs de succès et restitue les résultats pour les décideurs.

Définir la qualité ?

Dans une démarche de qualité, il est important de définir clairement les caractéristiques attendues ainsi que les critères d'évaluation de la qualité des données. Il est ensuite plus facile de mettre en œuvre les mesures de suivi et les plans d'actions de correction.

Données, informations et connaissances

Commençons par définir les concepts de donnée, d'information et de connaissance.

- une donnée est une description élémentaire, souvent codée, d'une chose, d'une transaction d'affaire, d'un événement, etc. Les données peuvent être conservées et classées sous différentes formes : papier, numérique, alphabétique, images, sons, etc.
- L'information représente les données transformées sous une forme significative pour la personne qui les reçoit : elle a une valeur pour ses décisions et ses actions
- Bien que la définition de la connaissance fasse encore débat parmi les philosophes, dans le monde de l'entreprise c'est le traitement des données et des informations qui permet de générer des connaissances : un moyen de compréhension ou d'apprentissage d'un problème ou d'une activité.

L'idée générale est de gérer les données comme un actif de l'entreprise au même titre que ses produits, ses employés, ses clients. Il faut donc comprendre les besoins des clients (ici les utilisateurs), créer des familles de données, c'est-à-dire toutes les données associées (Figure 3) et les gérer dans leur cycle de vie complet. On doit nommer un steward de données ayant un rôle similaire à un chef de produit.

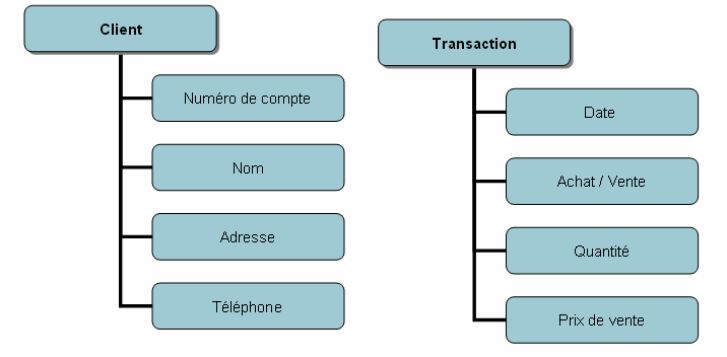


Figure 3 - Exemples de famille de données

Qualité des données

Une fois les données définies, nous pouvons expliciter ce qui fait leur qualité. C'est un terme générique décrivant à la fois les caractéristiques des données : complètes, fiables, pertinentes, à jour, cohérentes, mais aussi l'ensemble des processus qui permet de garantir ses caractéristiques. Le but est d'obtenir des données sans doublon, sans fautes d'orthographe, sans omission, sans variation superflue et conformes à la structure définie.
Les données sont dites de qualités si elles satisfont aux exigences de leurs utilisations. En d'autres termes, la qualité des données dépend autant de leur utilisation que de leur état. Pour satisfaire à l'utilisation prévue, les données doivent être exactes, opportunes et pertinentes, complètes, compréhensibles, et dignes de confiance. La Figure 4 illustre les nombreux aspects de la qualité des données.

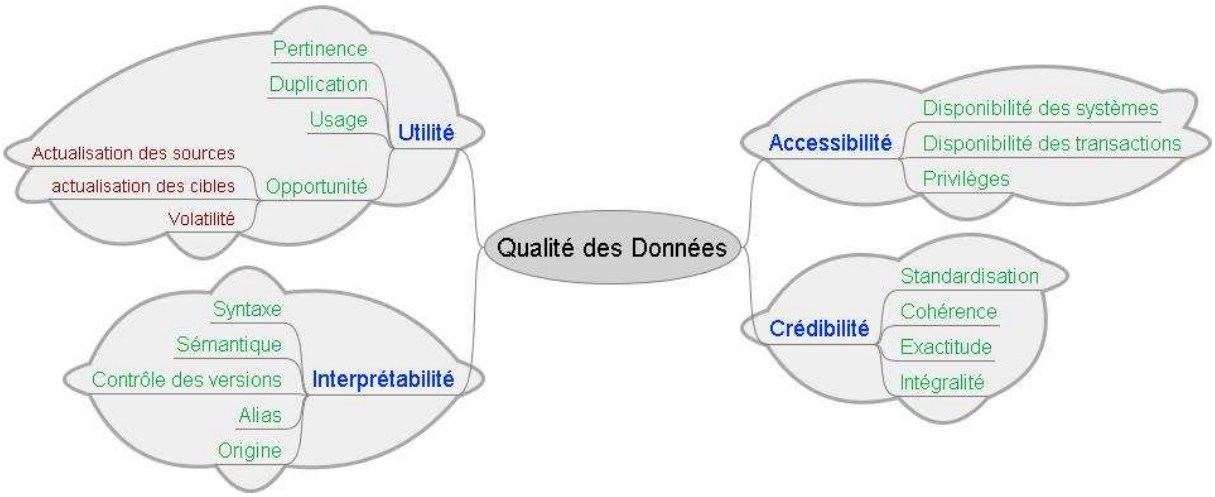


Figure 4 - Les dimensions de la qualité des données

Pour définir les problèmes de qualité dans votre entreprise, il est recommandé de définir les dimensions possibles et leur importance :

- Duplication : les données sont répétées. L'entité est gérée par plusieurs systèmes d'informations sous des identifiants différents et donc sa vue n'est pas unifiée.
- Standards : les valeurs sont correctes par rapport à un intervalle de répartition ou à un domaine. Par manque de standards de codification, l'entreprise « Les chantiers Techniques de Marseille » peut apparaître comme « Ets CTM », « C.T.M. » ou « CTM SA »
- Intégralité : toutes les données nécessaires sont disponibles pour le besoin métier. Il est impossible d'effectuer une campagne d'e-mailing avec une base de données clients ne contenant pas l'adresse email.
- Exactitude : les données représentent la réalité ou sont vérifiables à partir d'une source externe. Le code postal ne correspond pas à la localité, le téléphone a changé ou le SIRET n'a pas été mis à jour lors du déménagement de l'entreprise.
- Interprétabilité : une donnée doit être représentée sous un format cohérent et sans ambiguïté. Par exemple, affichée sous la forme 11/12/1963 sur l'écran du responsable du personnel de Paris, la date de naissance d'un employé est exacte, mais doit être affichée 12/11/1963 sur l'écran de son collègue américain.
- Opportunité : les données sont à jour au moment de leur utilisation. Le rapport mensuel des ventes doit inclure tous les résultats actualisés du mois pour toutes les régions commerciales.

Les données doivent avoir la qualité nécessaire pour supporter le type d'utilisation. En d'autres termes, la demande de qualité est aussi importante sur les données nécessaires à l'évaluation d'un risque que sur celles utilisées dans une opération de marketing de masse.

Indicateurs et Mesures

A partir de ces définitions théoriques, les organisations doivent créer leurs propres définitions opérationnelles en fonction des objectifs et priorités de l'entreprise, afin de définir les indicateurs pour chacune des dimensions, et vérifier par des mesures régulières leur évolution dans le temps.

Chaque dimension peut être mesurée soit de manière subjective en recueillant la perception des utilisateurs, soit de manière objective au travers de suivis automatiques des indicateurs spécifiques. La Figure 5 donne des exemples d'indicateurs de qualité suivant différents critères.

Critères de Qualité des Données	Caractéristiques	Exemples d'Indicateurs
Opportunité	L'âge des données est-il conforme aux besoins métiers ?	Date de la collecte des données Date du dernier traitement Contrôle de la version
Intégralité / Complétude	Est-ce que toutes les données nécessaires sont disponibles ?	Intégralité des valeurs optionnelles Nombre de valeurs non renseignées Nombre de valeurs par défaut par rapport à la moyenne
Cohérence	Quelles sont les données sources des informations contradictoires ?	Vérification de plausibilité Valeur de la déviation standard
Exactitude	Les valeurs représentent-elles la réalité ?	Fréquence des changements de valeur Réaction (feedback) des clients
Interprétabilité	Les données sont-elles compréhensibles par les utilisateurs ?	Valorisation des données utilisateur Violation de domaines
Standardisation, conformité	Quelles sont les données saisies, stockées ou affichées dans un format non standard ?	Certificat de conformité
Duplication	Quelles sont les données répétées ?	Nombre d'enregistrements dupliqués

Figure 5 - Exemples d'indicateurs de qualité

Une fois les indicateurs définis, il faut mettre en place un système de mesure qui permette de surveiller leur évolution dans le temps. La publication des indicateurs de qualité, leur cible et leur évolution permettent de définir les plans d'action éventuels à mettre en œuvre pour corriger une situation. La Figure 6 montre un exemple d'indicateur et son évolution dans le temps.

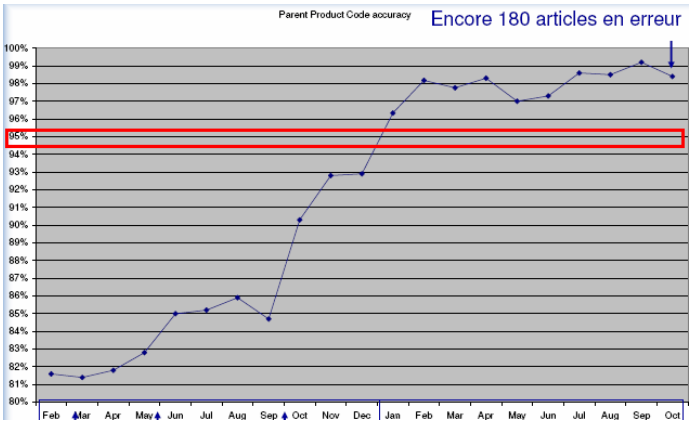


Figure 6 - Mesures d'un indicateur

Avec ces rapports, l'organisation est capable de déterminer les domaines d'amélioration et les plans d'actions associés, afin de remédier aux problèmes de qualité ainsi mis en évidence.

La gouvernance

Dans le cadre de l'initiative qualité de données, l'entreprise définit son modèle de gouvernance, c'est-à-dire son modèle de pilotage formalisé des personnes, processus et techniques pour faciliter la capacité à s'appuyer sur les données comme un atout majeur pour l'entreprise.

Rôles de la direction générale et des directions opérationnelles

Pour lancer cette démarche, deux garanties de succès doivent être réunies : le sponsoring de la direction générale, et l'implication de tous les acteurs. Il faut, pour convaincre la direction générale, prendre le temps de mesurer l'impact de la non-qualité et surtout démontrer que la qualité est source de compétitivité pour l'entreprise. Reste ensuite à faire preuve de pédagogie auprès des directions opérationnelles. Le directeur général ou le directeur des opérations, propriétaire des données, ne s'implique pas au quotidien dans la qualité des données. Cependant il doit s'assurer que l'initiative est lancée, et créer un comité Qualité des Données.

Le comité Qualité des Données

Le modèle de gouvernance doit comprendre une structure organisationnelle qui est chargée de l'amélioration de la qualité des données dans l'entreprise. Le comité Qualité des Données (Figure 7) est sous la responsabilité d'un sponsor, généralement nommé par un membre de la direction générale qui gère l'initiative. Le sponsor doit avoir une influence sur l'ensemble des directions métiers. Le comité a la responsabilité de la qualité des données de l'entreprise. Il définit les objectifs et priorités. Il s'assure que tous les projets incorporent la gestion de la qualité des données dans leurs processus de saisies, de transformations et de restitutions. Il s'assure également de la disponibilité des financements nécessaires à l'initiative. Il se réunit régulièrement pour assurer le suivi sur la qualité et faire le point sur les actions d'amélioration. Il décide des nouvelles priorités.

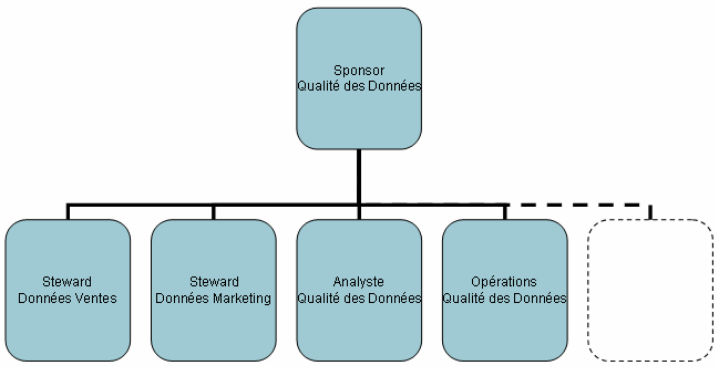


Figure 7 - Comité Qualité des Données

Ce comité est constitué d'experts issus des principales directions de l'entreprise, les stewards. Ces derniers sont responsables des données relevant de leur domaine d'expertise. Ils sont responsables de la définition et de la surveillance des mesures et indicateurs de qualités des données de leur domaine, et coordonnent les plans d'actions d'amélioration des indicateurs. L'analyste Qualité des Données est un professionnel de la DSI. Il met en application les règles métiers définies par les stewards dans les outils de profilage et de nettoyage.

Le socle technologique

Après avoir convaincu la direction générale et les directions métiers de l'importance de la qualité des données, après avoir mis en place la structure organisationnelle pour supporter l'initiative, il est temps d'évaluer les solutions technologiques. La mise en place d'une technologie de qualité des données doit permettre de :

- Faire les diagnostics et l'évaluation des problèmes de qualité
- Supporter les efforts d'intégration sur toutes les sources de données
- Automatiser le traitement des erreurs dans les processus d'extraction et de rechargement
- Définir un framework pour capturer et gérer l'ensemble des erreurs liées à la mauvaise qualité des données
- Procurer un cadre pour mesurer l'évolution des indicateurs qualité dans le temps
- Fournir des indicateurs de confiance sur la qualité des données utilisées

Les fonctions des outils

La plupart des solutions technologiques de qualité des données intègre des outils qui offrent les fonctions de qualité des données suivantes :

- Profilage (*Profiling*): analyse de la qualité des données afin de déterminer les domaines d'amélioration
- Standardisation : moteur de règles qui s'assure que les données sont conformes à des règles de qualité
- Nettoyage (*Cleansing*) : détection et correction des données corrompues ou inexactes
- Rapprochement (*Matching*) : comparaison et rapprochement des données pour découvrir des duplications éventuelles
- Enrichissement : utilisation de sources externes pour améliorer la complétude des données
- Décomposition (*Parsing*) : identification, vérification et décomposition un par un des éléments des zones de saisie libres
- Surveillance (*Monitoring*) : suivi de la qualité des données dans le temps et production de rapports de qualité

Les outils de **profilage** des données analysent l'état des données dans les bases de données ou les fichiers. Ils collectent des statistiques et des informations sur les données afin d'analyser si elles sont de qualité suffisante pour être utilisées dans d'autres contextes. Ils analysent la conformité des données par rapport aux standards de l'entreprise et aux définitions de ces champs (metadonnées). Ils identifient les dépendances avec les autres sources de données et évaluent les duplications d'information.

En utilisant les règles définies par les métiers, les outils de **standardisation** et de **validation** automatisent le processus de vérification et de correction des données afin que les abréviations soient standardisées, les données correctement orthographiées et les modèles de formatage correctement utilisés. Ils valident les valeurs des données par rapport à un intervalle de répartition ou à un domaine (par exemple : validation des adresses suivant les standards postaux).

Les outils de **nettoyage** permettent de détecter et de corriger (ou de supprimer) des enregistrements corrompus ou inexacts d'une base de données ou d'un fichier. Les erreurs détectées ont pu être créées dans des environnements applicatifs hétérogènes, saisies en erreur par un utilisateur ou corrompues lors d'une transmission ou du stockage. L'objectif du nettoyage est de rendre la source de données cohérente avec les autres sources de l'entreprise. Les outils de nettoyage sont utilisés a

posteriori sur les données, à la différence des outils de standardisation et de validation qui sont utilisés lors de la saisie des données.

Les outils de **rapprochement** permettent de comparer des données de sources différentes. Ils permettent d'identifier les relations entre les enregistrements de données afin de les dédupliquer ou de réaliser des traitements par groupe. Ils permettent d'identifier les enregistrements qui décrivent la même entité.

Les outils de **décomposition** permettent de transformer un champ de saisie contenant des données multiples dans une structure généralement arborescente utilisée par les applications. Par exemple, les outils de parsing peuvent être utilisés pour reconnaître dans un champ les données d'adresses, des mesures, des quantités ou des références produits.

De même, les outils **d'enrichissement** permettent d'ajouter à des enregistrements, des données en provenance d'autres sources internes ou externes.

Enfin, les outils de **surveillance** permettent d'identifier et de réagir immédiatement aux problèmes avant que la qualité des données ne se dégrade. Ils permettent de suivre l'évolution des données dans le temps et de déterminer leur détérioration éventuelle. Ils identifient les tendances sur la qualité des données et alertent sur les violations des règles de qualité définies.

Il est clair que ces différents outils qui gèrent les différents aspects de la qualité des données ne sont pas indépendants les uns des autres. La Figure 8 illustre l'imbrication des différentes étapes qui amènent à une vue unique des informations Client dans un processus bancaire.

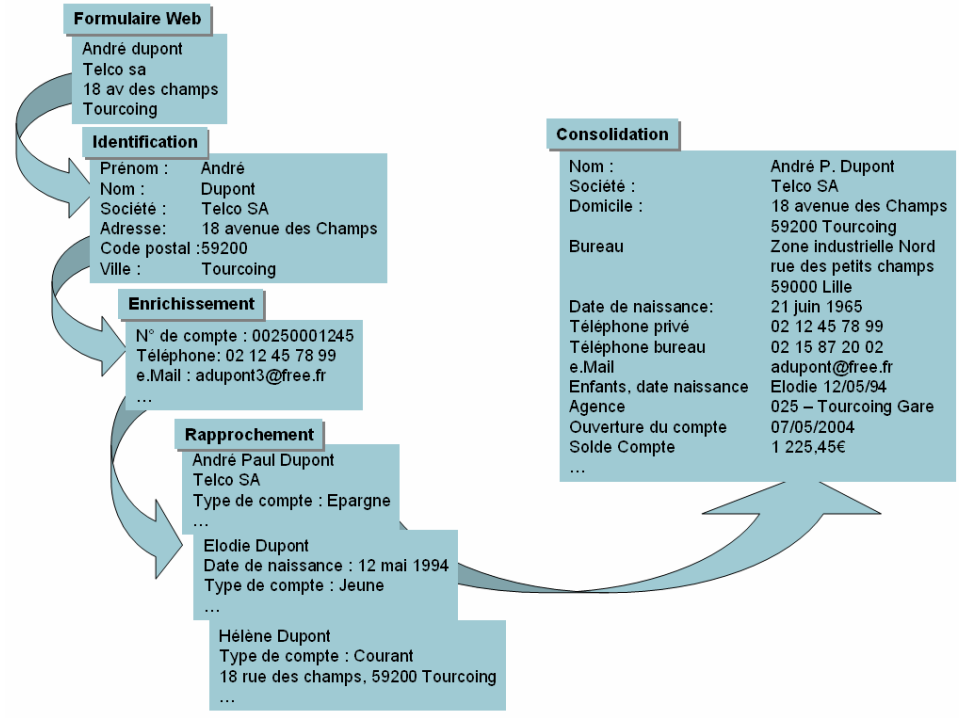


Figure 8 - Un processus de qualité

Une infrastructure de qualité

Les services de qualité des données sont au centre de l'infrastructure informatique. L'entreprise doit maintenant définir son architecture de données et l'infrastructure technique associée, en particulier l'ensemble des services qui garantissent leur qualité. Au-delà des services d'accès et d'intégration, il faut maintenant mettre en œuvre des services de qualité lors de la saisie, du traitement et de la restitution des données. La Figure 9 décrit le rôle central des services de qualité dans l'architecture globale de données.

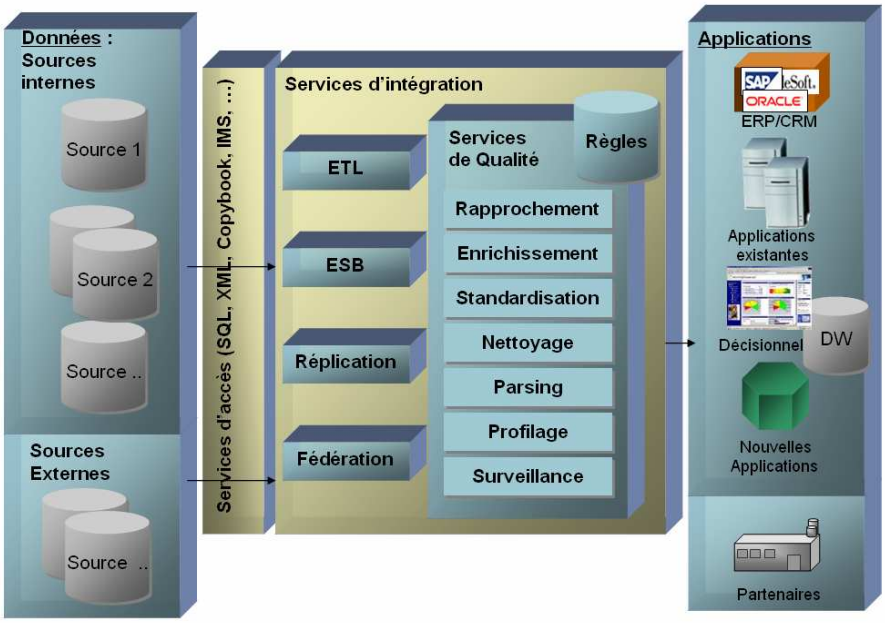


Figure 9 - Les services de qualité des données

Au-delà de la vision produit, on doit passer d'une logique de projet à une logique d'entreprise. Cette vision de plate-forme de service doit définir les briques logicielles pour répondre aux besoins métiers de qualité des données. Du point de vue technique, il faut s'assurer que l'implémentation permette l'interopérabilité avec les outils d'intégration de données (ETL, EII, ESB…), ainsi qu'avec l'ensemble des applications métiers, patrimoniales, les data warehouses et autres data marts.

Scénarios de mise en œuvre

La qualité des données a souvent été analysée de manière isolée pour répondre à la problématique d'un département ou localement sur une base de données. Historiquement, les outils ont été déployés dans les silos applicatifs en mode batch comme une étape d'un processus déconnecté. Ce manque de coordination et l'éclatement des systèmes d'information renforce les risques de duplication, augmente la difficulté de mise à jour, génère une perte de contrôle de conformité. Pourtant les utilisateurs travaillent sur des données en provenance de sources multiples dans ces systèmes et applications distribués. Les directions métiers demandent que toutes les données de l'entreprise soient disponibles, accessibles, réutilisables et à jour. Les nouveaux projets informatiques exploitent des données collectées pour un objectif particulier dans des applications d'un autre domaine comme la Business Intelligence ou la gestion de la relation client.

Cette omniprésence des données au cœur de tous les domaines d'activité indique clairement que l'amélioration de leur qualité doit être un processus permanent répondant à un besoin global de l'entreprise. La gestion de la qualité des données fait donc partie des projets structurants de l'entreprise. Que ce soient l'efficacité des départements marketing et commercial par la mise en place d'un data warehouse, l'amélioration des performances opérationnelles d'un projet de gestion des données de référence (Master Data Management), en passant par l'optimisation du service aux clients par l'installation d'un outil de GRC (CRM) ou encore la nécessité de fournir des informations transparentes dans le respect des contraintes réglementaires, toutes ces initiatives nécessitent des données fiables et cohérentes entre elles, et par conséquent, de qualité.

Pour démarrer cette initiative, étant donné que les données de qualité sont au cœur de tous les grands projets stratégiques de l'entreprise, il est donc d'autant plus recommandé de la lancer simultanément à un projet stratégique. Cela permet de justifier l'analyse globale des données de l'entreprise et d'affecter en priorité des projets de résolution des problèmes identifiés, en cohérence avec la stratégie métier. L'initiative permet ainsi de réaliser le projet sur des fondations solides : des données de qualité.

Plusieurs domaines stratégiques sont dépendants de la mise à disposition et du maintien de données de qualité :

- La business intelligence
- La conformité réglementaire
- Les données de référence (master data)
- Le service aux clients
- La consolidation et l'intégration de données

Business Intelligence & Data Warehouse

L'informatique décisionnelle (BI pour Business Intelligence) désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données d'une entreprise, afin d'offrir une aide à la décision et de permettre aux responsables de la stratégie d'avoir une vue d'ensemble de l'activité traitée.
Ce type d'application utilise en règle générale un entrepôt de données (data warehouse) pour stocker des données transverses provenant de plusieurs sources hétérogènes et fait appel à des traitements lourds de type batch pour la collecte de ces informations.
Les applications classiques « d'entreprise » permettent de stocker, restituer, modifier les données des différents départements opérationnels de l'entreprise (production, marketing, facturation comptabilité, etc.). Ces départements possèdent chacun une ou plusieurs applications propres, et les données y sont rarement structurées ou codifiées de la même manière que dans les autres départements.

Chacun dispose le plus souvent de ses propres tableaux de bord et il est rare que les indicateurs (par exemple : le chiffre d'affaires sur un segment précis de clientèle) soient mesurés partout de la même manière, selon les mêmes règles et sur le même périmètre.
Pour pouvoir obtenir une vision synthétique de chaque service ou de l'ensemble de l'entreprise, il convient donc que ces données soient filtrées, croisées et reclassées dans un entrepôt de données central. Cet entrepôt de données va permettre aux responsables de l'entreprise et aux analystes de prendre connaissance des données à un niveau global et ainsi de prendre des décisions plus pertinentes.
Il est clair que l'initiative Qualité de Données trouve toute sa place dans le projet de Data Warehouse. D'une manière générale, la qualité des données est de la responsabilité de l'équipe du projet. En effet, cette dernière doit préparer pour les utilisateurs, des données exploitables et donc de qualité, le succès du projet en dépendant. La qualité des données permet tout d'abord de filtrer les données pour ne conserver que les *bonnes* données. Par ses processus de contrôle, elle permet également de valider que le projet est bien en ligne avec les besoins des utilisateurs en mettant à disposition des données bien choisies, accessibles, complètes et en temps utile. Par la mise en place d'indicateurs et métriques associés, elle permet enfin de vérifier que les utilisateurs comprennent la structure de l'entrepôt et qu'ils peuvent accéder aux données facilement.

Conformité réglementaire

La gouvernance d'entreprise et les questions de transparence de l'information financière sont au centre des débats depuis quelques années en France. L'éclatement de la bulle financière a déstabilisé les marchés et entâché la confiance des investisseurs. De plus en plus de données, généralement très dynamiques, issues de nombreuses applications sources, sont utilisées pour gérer les processus d'analyse de risques et de conformité réglementaire. Un ensemble de normes et de réglementations dont le nombre s'accroit régulièrement contrôle l'activité des entreprises. Il oblige les directions générales et leurs directeurs financiers à envisager les notions de risque et de conformité réglementaire sous un aspect global et dans le cadre général de l'entreprise.

Ces nouvelles exigences placent les entreprises dans l'obligation d'analyser en détail leur « chaîne d'information ». Elles doivent être en mesure de tracer l'information émise et de remonter la chaîne pour identifier les données et les décisions prises à partir de l'information.

Ici encore, l'initiative Qualité des Données a un rôle primordial dans la gestion des risques et la conformité réglementaire. Il semble difficile de gérer la conformité réglementaire sans faire confiance aux données. De plus, consolider des informations nécessite que chaque entité partage les mêmes définitions. Au-delà de l'aspect obligatoire – les entreprises n'ont pas d'autre choix que d'appliquer lois et règlements – il s'agit d'exploiter ce besoin de conformité pour améliorer la profitabilité et l'environnement de prise de décision. L'initiative permet de gérer d'une manière intégrée et globale la qualité des données. Elle donne les moyens de mesurer et surveiller cette qualité. Par exemple, l'évaluation de la qualité des données fait partie de la directive Bâle II. La démonstration du niveau de qualité des données par un processus documenté est certainement un point de contrôle externe appréciable. Enfin, l'initiative Qualité des Données permet d'agir sur les domaines identifiés d'améliorations sans impacter la qualité actuelle.

Bâle II

Les règles de transparence imposées par les directives européennes Bâle II implique de mettre en place un processus de consolidation des vues risques, comptables et financières des données qui doit reposer sur des données « dignes de confiance ».
L'organisme financier devra prouver a posteriori la validité de ses méthodes définies a priori, en fonction de ses données statistiques et cela sur des périodes assez longues (5 à 7 ans). Elle devra en outre être capable de "tracer" l'origine de ses données.
Suivant le même canevas, de nouvelles normes Solvabilité II sont en cours de discussion pour les sociétés d'assurance et de réassurance.

Données de référence (Master Data)

Aujourd'hui les frontières entre les services, les canaux de distributions et les départements des organisations disparaissent. Il s'agit d'optimiser la participation et les interactions entre tous les services au-delà des frontières administratives ou techniques. Tous les acteurs doivent partager un langage commun autour des entités gérées par l'entreprise : ses clients, ses produits, ses entités légales, ses employés, etc. De même, les équipes informatiques tentent de réduire l'impact des silos isolants les différents systèmes applicatifs (ERP, CRM, SCM, etc.). On assiste aujourd'hui à une demande croissante, tant du côté des métiers que du côté de l'informatique, de la création et la gestion d'un ensemble de données de référence (Master Data Management – MDM). L'objectif d'un projet de MDM est d'offrir à l'organisation une vue unique et unifiée des données à partir des multiples applications opérationnelles.
L'initiative Qualité des Données a un rôle fondamental dans le projet de MDM. Elle permet de standardiser, vérifier et éventuellement corriger les données en provenance de multiples sources opérationnelles. Elle permet également de faire les rapprochements de différents éléments des entités et de résoudre la duplication des données sur les clients et les produits dans un enregistrement de référence. Les projets de MDM se focalisent généralement sur l'accès et la transmission des données. Avec le support de l'initiative Qualité des Données, le projet MDM offre de meilleures données dans le référentiel général et par delà, une meilleure image de la réalité de l'entreprise.

Service aux clients

Afin de développer une stratégie centrée sur ses clients qui valorise, fidélise et personnalise les relations, l'entreprise doit disposer de toutes les informations clients et maîtriser toutes les interactions avec ces derniers. Les outils de gestion de la relation clients (Customer Relationship Management – CRM), les applications marketing, les centres d'appels permettent de créer et entretenir une relation mutuellement bénéfique entre l'entreprise et ses clients. La faible qualité des données liées aux relations entre l'entreprise et ses clients peut compromettre la rentabilité de l'investissement des projets de services aux clients, voire de détériorer les relations.
Il est critique d'assembler, présenter et maintenir des données lors de toutes les interactions avec les clients, depuis l'orthographe correcte du nom du client jusqu'aux mises à jour dynamiques des listes de prix sur le site Web de l'entreprise. Ici encore, le dynamisme de la base de données n'est pas à négliger. Conserver les données clients exactes dans le temps est un défi important. Il est impératif de mettre en œuvre un programme de gestion de la qualité des données clients qui évalue et met en place des processus de maintenance permanente de cette qualité (conversion, formatage, nettoyage, déduplication, etc.).

Consolidation et intégration

Le succès d'une fusion ou d'une acquisition réside en grande partie dans la rapidité d'unification et d'assimilation des deux organisations dans une entité unique. Le facteur temps est en effet un élément critique dans l'évaluation du retour sur investissement de la fusion. En parallèle et en support de l'organisation, la direction informatique est au défi d'intégrer les systèmes et applications de chaque entreprise rapidement. Ici encore, la qualité des données est sur le chemin critique de cette unification. Il faut pouvoir homogénéiser les sources, permettre l'échange et l'intégration des données entre les deux entités et garantir l'accès à des données standards pour toutes les fonctions et directions. Il faut aussi permettre d'obtenir une vue unique des données de référence pour toutes les entités opérationnelles et les différentes filiales à l'étranger.

De la même manière, l'exploitation des applications patrimoniales dans les nouvelles architectures, la consolidation des environnements pour réduire les coûts d'exploitation, le partage de données au-delà des frontières de l'entreprise nécessaire pour intégrer et automatiser la chaîne d'approvisionnement, rendent la qualité des données incontournable. Des données de mauvaise qualité affectent la performance globale des professionnels. Les responsables ne font pas confiance aux informations

présentées. Les utilisateurs passent du temps à corriger les erreurs manuellement. De plus, les coûts d'exploitation et les risques liés aux projets augmentent fortement, alors que les délais s'allongent pour traiter les corrections. Au contraire, travailler sur des données de qualité permet de faciliter les opérations et d'optimiser l'usage de ressources en général limitées. Les équipes informatiques peuvent se consacrer aux développements d'applications innovantes et à la maintenance des données, plutôt qu'à la détection et la résolution d'anomalies.

L'offre Qualité de Données d'Informatica

Informatica Corporation est un éditeur de solutions d'intégration et de qualité de données d'entreprise qui permettent aux organisations d'accéder, intégrer, migrer et consolider les données générées et utilisées par l'ensemble de leurs systèmes, processus et collaborateurs. Ces solutions permettent de réduire la complexité, garantir la cohérence et accroître la performance globale des activités des entreprises. Acteur historique sur ce marché, Informatica a complété son offre sur la qualité des données, à la suite du rachat de la société Similarity Systems en janvier 2006 qui avait elle-même acquis les actifs de la société Evoke Software fin 2005. Informatica propose aujourd'hui l'ensemble des services de qualité des données grâce à ses produits Informatica Data Explorer et Informatica Data Quality.

Les offres de qualité des données complètent les offres traditionnelles d'Informatica :

- Informatica PowerCenter la plate-forme d'extraction, transformation et chargement (Extract, Transform, Load – ETL). Elle permet aux organisations d'accéder et d'intégrer des données à partir de presque tout système d'entreprise, et ce, quel qu'en soit le format, puis de les transmettre à toute l'entreprise au moment voulu.
- Informatica PowerExchange fournit « à la demande » un accès immédiat à tous les systèmes de données critiques de l'entreprise—mainframes, bases de données relationnelles, systèmes à base de fichiers, etc.

Informatica Data Explorer et **Informatica Data Quality** apportent des capacités de diagnostic des données afin de comprendre, d'identifier, et de localiser les problèmes pour mieux en qualifier les incohérences et en assurer la correction.
Informatica Data Explorer se concentre sur les tâches de profilage des données et ses résultats peuvent alimenter un processus d'intégration. Il analyse les données et produit un modèle complètement normalisé des données.
Informatica Data Quality est un outil d'analyse, de nettoyage, de correction et de déduplication de données. Il permet d'identifier et de résoudre tout type de problème de qualité de données, afin de les préparer pour une consolidation ou un processus de chargement.

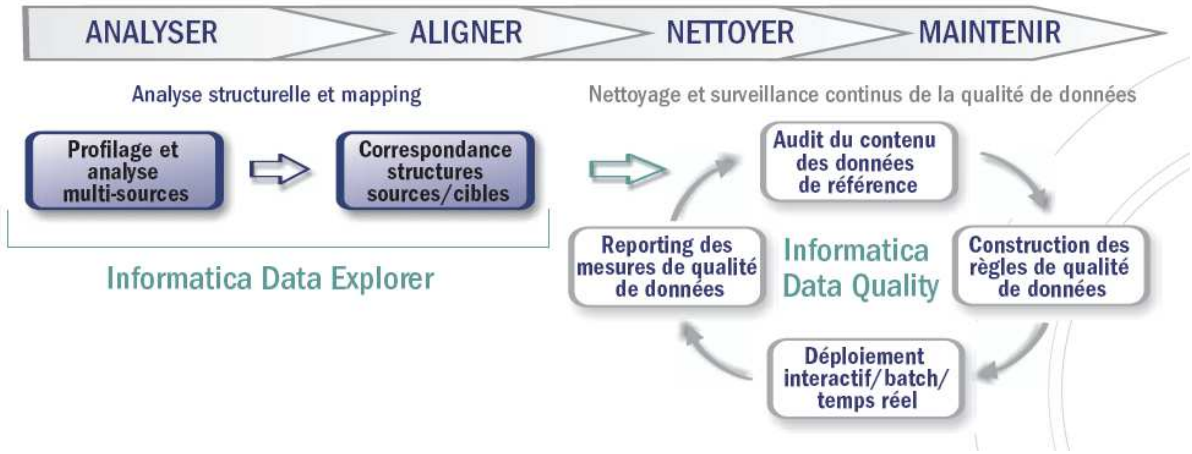


Figure 10 - Le processus de gestion de la qualité des données d'Informatica

Informatica Data Explorer

Informatica Data Explorer permet une évaluation et un profilage approfondis de multiples sources de données. Informatica Data Explorer accède aux principaux types de sources de données, notamment fichiers plats, bases de données relationnelles, mainframe VSAM et IMS.

Analyser

La qualité des données commence par la compréhension de toutes les données qui se trouvent dans les systèmes sources. Informatica Data Explorer permet de découvrir automatiquement et d'analyser les données afin de corriger les problèmes de qualité. Informatica Data Explorer met en œuvre un processus de profilage qui consiste à identifier le contenu, la structure et la qualité réelle de l'ensemble des données. Ce processus est effectué selon trois dimensions : les valeurs des attributs de chaque table, les relations entre les attributs de chaque table, et les données entre les tables pour découvrir les attributs identiques ou se recoupant.

Aligner

A partir des informations collectées lors du processus de profilage des données, Informatica Data Explorer construit un modèle de données tiers normalisé dans lequel les redondances non souhaitées sont éliminées. Ce modèle peut alors être utilisé comme zone intermédiaire pour déplacer des données vers une cible déterminée ou en tant qu'entrepôt de données opérationnelles. L'ensemble des informations découvertes sont stockées dans un référentiel. Elles sont disponibles aux processus de nettoyage (Informatica Data Quality) et d'intégration (Informatica PowerCenter).

Informatica Data Quality

Informatica Data Quality fournit aux analystes métiers une plate-forme pour concevoir, gérer et déployer des processus de qualité de données.

Nettoyer

Avec le profil des sources de données, Informatica Data Quality peut mettre en œuvre un processus automatisé de nettoyage des données. Le *Designer* permet de créer des règles, normes et données de référence relatives à la qualité de données et de les déployer à l'ensemble de l'entreprise. Il permet également de gérer des tableaux de bord permettant la mesure et la surveillance d'indicateurs clés de qualité. Le *Runtime* et le *Realtime* permettent de déployer ces programmes de qualité de données en mode batch ou temps réel.

Maintenir

Informatica Data Quality permet une amélioration continue de la qualité des données au travers de son processus de management itératif et de ses fonctions de création de rapports. Il est conçu pour être utilisé par les équipes d'analystes de données et les stewards. La solution fournit des tableaux de bord qui assurent la surveillance des principaux paramètres de la qualité de données (complétude, conformité, cohérence, exactitude, duplication et intégrité) pour toutes les données. Les rapports permettent aux utilisateurs de descendre à des niveaux de détail plus fins pour examiner les enregistrements de mauvaise qualité un par un, et identifier les problèmes dans le cadre d'un processus itératif de découverte et de nettoyage.

Services

Méthodologie

Le recueil de bonnes pratiques **Informatica Velocity** offre un cadre d'implémentation des solutions d'Informatica et en particulier des solutions de qualité des données. Il couvre les principales phases du projet de qualité de données.

Offres de services

Bien entendu, Informatica dispose d'un programme d'assistance complet pour le déploiement de ses solutions. En particulier, Informatica propose les offres suivantes dans le domaine de la qualité des données :

- Déploiement d'Informatica Data Quality
- Déploiement de Data Cleanse et Match
- Audit d'Informatica Data Quality
- Quick start de l'option Web services d'Informatica Data Quality
- Quick start de l'option rapports et tableaux de bord d'Informatica Data Quality

Conclusion

En conclusion, l'amélioration de la qualité des données de l'entreprise passe par la mise en place d'une initiative continue et globale.

Ce livre blanc a évoqué les concepts de qualité des données, son importance dans les organisations, entreprises grandes ou petites et administrations. Une mauvaise qualité des données coûte cher et conduit à des ruptures dans les processus, à des décisions métiers moins pertinentes et à une gestion médiocre de la relation client. De plus, elle peut invalider les efforts de l'entreprise en matière de conformité réglementaire.

Il est recommandé de s'adosser à un grand projet stratégique dans l'entreprise pour lancer une initiative autour de la qualité des données. Mais, cette initiative peut aussi être menée de façon indépendante. L'idée de démarche et de pérennité est essentielle et caractéristique de l'approche qualité. Elle va à l'encontre d'une opération unique et isolée qui ne permet de nettoyer et d'améliorer les données que ponctuellement. Cela signifie que les objectifs, mesures et indicateurs associés doivent être portés par l'ensemble des acteurs concernés et notamment une implication forte de la hiérarchie.

Cette démarche doit être lancée conjointement par les directions métiers, pour leurs connaissances des impératifis liés à leur activité et des stratégies de l'entreprise, et la direction informatique, pour son expertise technologique. Elle passe d'abord par la connaissance de données. Il est nécessaire d'évaluer l'état des données de votre organisation avec un focus sur l'aspect de qualité. L'aspect organisationnel est important. Il est crucial de construire une équipe mixte métier et informatique, le comité Qualité des Données, ayant les compétences nécessaires et du temps disponible pour s'atteler à cette tâche. Cette équipe aura pour mission de définir les principaux indicateurs et mesures de la qualité des données, justifier les programmes d'amélioration à mettre en œuvre et de mesurer de façon régulière les progrès effectués.

La technologie permet d'automatiser les tâches de contrôle et de nettoyage, ainsi que la production des indicateurs et des rapports. Elle supporte d'une manière efficace les demandes des directions métiers. Elle prépare, transforme et propose les informations clés de prise de décision. Mais la technologie n'est qu'un élément de la solution. Les ordinateurs gèrent les données, les utilisateurs exploitent les connaissances.

La qualité des données est avant tout un problème métier, pas seulement un problème informatique. Plus elle sera incorporée aux habitudes et à la culture de l'entreprise, plus la démarche qualité progressera. Paradoxalement, son succès résidera dans sa banalisation.

A propos de JEMM research et de l'auteur:

JEMM research est une société de recherches stratégiques et d'analyses opérationnelles, spécialisée dans les infrastructures logicielles, les systèmes ouverts, et les approches orientées services. JEMM research conseille les entreprises utilisatrices sur l'évolution de leur organisation, dans leur choix d'architecture et de technologies, les aide dans les étapes du projet d'évolution de leur système d'information, les accompagne dans le changement, et valide et documente les réalisations. Par ailleurs, JEMM research aide les éditeurs de logiciels à comprendre, analyser, définir leurs marchés cibles et à promouvoir leur offres en maximisant leur chances de succès.

Christophe TOULEMONDE est Directeur du cabinet JEMM research. Avec plus de 20 ans d'expérience dans l'informatique, Christophe est un expert reconnu des architectures orientées-services, spécialiste de l'infrastructure et de l'intégration d'entreprise (données, applications, processus), du design et de l'architecture des applications distribuées et plus généralement de l'architecture d'entreprise.
Auparavant, chez Meta Group, il a couvert, pour la zone EMEA, les domaines des stratégies d'intégration et de développement. Pendant 15 ans chez IBM et des filiales du groupe en France et aux Etats-Unis, il a occupé divers postes de direction technique et marketing. Il a publié de nombreux ouvrages sur le e-business et l'intégration d'applications.

JEMM research

www.jemmresearch.com

jemminfo@jemmresearch.com
Tel : +33 1 39 16 48 81

INFORMATICA
The Data Integration Company™

Informatica France : Immeuble Le Linéa, rue du Général Leclerc -
92047 Paris La Défense Cedex (France)
Tél. : + 33 1 41 38 92 00 – Fax : + 33 1 41 38 92 01 – www.informatica.com/fr

Informatica Division Data Quality : Wilson House, Fenian Street – Dublin 2 – Irlande
Tél. : +353 1 4004900 - Fax : +353 1 4004999 - www.informatica.com

Siège international : 100 Cardinal Way, Redwood City, CA 94063 (USA)
Tél. : + 1 650 385 5000 Fax : + 1 650 385 5500 N° Vert USA : + 1 800 970 1179
www.informatica.com

Informatica dans le monde : Allemagne • Australie • Belgique • Canada • Etats-Unis • France • Japon • Pays-Bas • Royaume-Uni • Singapour • Suisse