

Cinq pratiques fondamentales en matière de qualité des données

WHITE PAPER:
QUALITÉ DES DONNÉES ET INTÉGRATION DES DONNÉES

David Loshin



Cinq pratiques fondamentales en matière de qualité des données

Introduction

LA GESTION DE LA QUALITÉ DES DONNÉES CONSTITUE UN « CYCLE VERTUEUX » DANS LEQUEL L'ANALYSE, L'OBSERVATION ET L'AMÉLIORATION CONTINUES ENTRAÎNENT UNE AMÉLIORATION GLOBALE DE LA QUALITÉ DES INFORMATIONS ORGANISATIONNELLES. LES RÉSULTATS DE CHAQUE ITÉRATION PEUVENT AMÉLIORER L'ACTIF INFORMATIONNEL D'UNE ENTREPRISE AINSI QUE LA FAÇON DONT IL PERMET D'ATTEINDRE LES OBJECTIFS OPÉRATIONNELS.

CE CYCLE S'ARTICULE AUTOUR DE L'EXÉCUTION DE CINQ PRATIQUES FONDAMENTALES EN MATIÈRE DE GESTION DE LA QUALITÉ DES DONNÉES, QUI SONT MISES EN ŒUVRE AU MOYEN D'UNE COMBINAISON DES SERVICES DE DONNÉES ESSENTIELS. CES PRATIQUES SONT LES SUIVANTES :

- ÉVALUATION DE LA QUALITÉ DES DONNÉES
- MESURE DE LA QUALITÉ DES DONNÉES
- INTÉGRATION DE LA QUALITÉ DES DONNÉES DANS L'INFRASTRUCTURE DES APPLICATIONS
- AMÉLIORATION DE LA QUALITÉ DES DONNÉES OPÉRATIONNELLES
- GESTION DES INCIDENTS AU NIVEAU DE LA QUALITÉ DES DONNÉES

EN METTANT EN PLACE DES PROCESSUS REPRODUCTIBLES EN MATIÈRE DE GESTION DE L'OBSERVATION DES ATTENTES DANS LE DOMAINE DE QUALITÉ DES DONNÉES, CES PRATIQUES PROCURENT UN SUPPORT SOLIDE À LA GESTION DE LA QUALITÉ DES DONNÉES DE L'ENTREPRISE. CE DOCUMENT PRÉSENTE CES PRATIQUES PUIS EXAMINE LES SERVICES DE DONNÉES SOUS JACENTS CHAQUE SECTION PROPOSE UN PANORAMA DE LA PRATIQUE ET ÉTUDIE LES PROCESSUS INTERVENANT DANS L'ATTEINTE DES OBJECTIFS. DES PRATIQUES SAINES DE GESTION DES DONNÉES ALLIÉES À UNE PLATEFORME TECHNOLOGIQUE ADAPTÉE PERMETTENT À UNE ENTREPRISE D'INCORPORER LA QUALITÉ DES DONNÉES À L'ARCHITECTURE EN PLACE.

LES PRATIQUES DE GESTION DE LA QUALITÉ DE DONNÉES SONT EN MESURE D'AMÉLIORER LA VALEUR DE L'ACTIF INFORMATIONNEL D'UNE ENTREPRISE AINSI QUE LA FAÇON DONT IL SUPPORTE LA RÉALISATION DES OBJECTIFS OPÉRATIONNELS.

Le cycle vertueux de la qualité des données

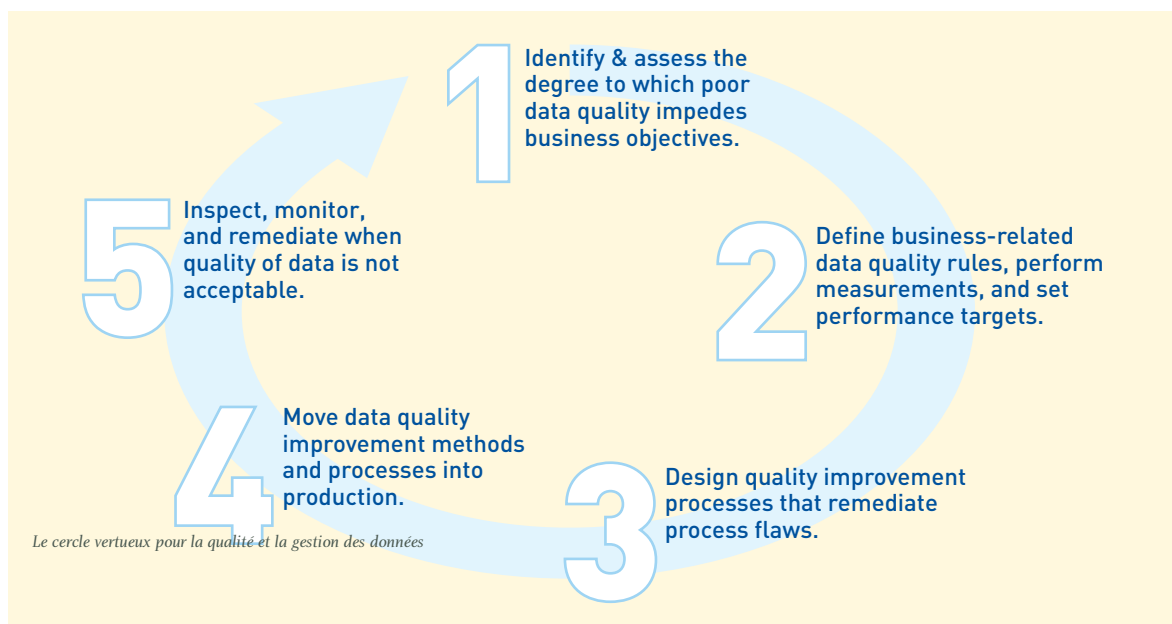
La gestion de la qualité des données constitue un « cycle vertueux » dans lequel l'analyse, l'observation et l'amélioration continues entraînent une amélioration globale de la qualité des informations organisationnelles. L'objectif de ce cycle est de passer d'une entreprise dans laquelle les gestionnaires de données réagissent face aux problèmes de données à une entreprise qui maîtrise et limite, de manière proactive, l'introduction de données erronées dans son environnement.

Ce cycle englobe cinq pratiques fondamentales en matière de gestion de la qualité des données, qui sont mises en œuvre au moyen d'une combinaison des services de données essentiels. Ces pratiques sont les suivantes :

- 1 Évaluation de la qualité des données, qui permet au spécialiste de comprendre l'incidence de la médiocrité des données sur les processus de gestion et la façon dont ils sont censés s'exécuter ainsi que d'élaborer un dossier d'analyse pour la gestion de la qualité des données.
- 2 Mesure de la qualité des données, dans le cadre de laquelle l'analyste de la qualité synthétise l'évaluation

des résultats et se concentre sur les éléments de données considérés comme étant essentiels en fonction des besoins de l'utilisateur. Cela permet de définir la métrologie des performances qui alimente les rapports de gestion par le biais de cartes de notation.

- 3 Intégration de la qualité des données dans l'infrastructure des applications, en intégrant l'analyse des besoins en matière de données dans toute l'entreprise et en incorporant la qualité des données dans le cycle de vie du développement.
- 4 Amélioration de la qualité des données opérationnelles, dans laquelle les procédures de gestion des données servent à gérer les règles de qualité des données identifiées ainsi que la conformité aux seuils d'acceptabilité.
- 5 Gestion des incidents de la qualité des données, qui permet aux analystes de la qualité d'étudier si les données respectent les niveaux d'acceptabilité, de signaler, consigner et effectuer un suivi des problèmes ainsi que de documenter les processus de correction et de perfectionnement.



Cinq pratiques fondamentales en matière de qualité des données

Collectivement, ces pratiques établissent le fondement du programme de gestion de la qualité des données puisqu'elles donnent lieu à un processus reproductible en vertu duquel des paramètres contribuent au remplissage d'une carte de notation et d'un tableau de bord ainsi qu'à la mise en place d'une gestion proactive de la qualité des données. Les employés formés doivent avoir recours à ces services de données essentiels pour que ces pratiques deviennent opérationnelles.

Évaluation de la qualité des données

Les entreprises intelligentes veulent maximiser l'investissement qu'elles réalisent au niveau de la gestion de la qualité des données. Cela signifie comprendre l'incidence négative que peut avoir une qualité de données médiocre sur l'atteinte des objectifs opérationnels. En quantifiant cet écart, le spécialiste est en mesure de déterminer le rapport coût-efficacité, la faisabilité et la vitesse des propositions d'amélioration de la qualité des données. La compréhension de l'impact des flux de données dans un contexte de gestion procure un critère d'évaluation et de hiérarchisation des problèmes de données émergents.

À titre d'exemple, il se pourrait que les coûts d'affranchissement et d'expédition soient surévalués si les adresses indiquées sont incorrectes. Cette idée de hausse de coûts pourrait être introduite du fait du volume important de colis impossibles à livrer et donc renvoyés. Toutefois, les adresses incorrectes n'entraînent pas seulement la hausse des coûts directs associés aux articles renvoyés. En effet, les applications analytiques servant au profilage des tendances d'achat des clients par région seront également erronées, ce qui pourrait affecter la bonne exécution des campagnes commerciales et des promotions régionales. Le processus d'évaluation de la qualité des données peut servir à quantifier ces coûts et conséquences ainsi qu'à déterminer le pourcentage pouvant être directement lié aux adresses rectifiables.

Ces pratiques incorporent les processus permettant d'identifier, d'évaluer, de quantifier et de hiérarchiser les problèmes de qualité :

- Analyse des conséquences opérationnelles – Ce processus a pour objectif de guider les analystes en signalant tous les problèmes de données potentiels susceptibles d'augmenter les coûts, de réduire les revenus, d'influencer les marges ou d'introduire des retards dans les activités. Essentiellement, l'objectif est d'identifier toute incidence négative pouvant être attribuée aux données dont la qualité est inacceptable. L'identification de l'emplacement et de la magnitude des points névralgiques critiques figurant dans les différents processus de gestion permet de déterminer les besoins en informations dans le cadre de l'évaluation, de restreindre la liste des jeux de données à examiner et de guider l'identification des exigences en matière de qualité des données.
- Évaluation de la qualité des données au moyen du profilage – Ce processus réalise un examen ascendant des données réelles afin d'isoler les anomalies apparentes susceptibles de représenter des défauts véritables. Au moyen du profilage des données et autres techniques statistiques et analytiques, les analystes sont en mesure d'identifier ces anomalies apparentes, anomalies qui pourront être étudiées plus en avant avec l'examen par les utilisateurs des données opérationnelles.
- Revue des anomalies soulignées par l'évaluation de la qualité des données – Au cours de ce processus, les analystes examinent les anomalies mises en lumière avec les consommateurs de données opérationnelles pour constater si un lien existe entre les erreurs de données et les incidences opérationnelles éventuelles. En distinguant les erreurs ayant un impact matériel de celles sans pertinence, l'équipe est en mesure de hiérarchiser les problèmes en fonction de leur incidence opérationnelle ainsi que d'explorer les différentes façons de les résoudre.
- Définir les mesures de la qualité des données – La corrélation des incidences opérationnelles et des problèmes de données par le biais de règles définies permet d'élaborer une méthode de mesure destinée à référencer les niveaux de qualité ainsi qu'à mener à bien une observation et inspection continues au sein du flux de production des informations. Ce processus guide la prise en compte des mesures à adopter ainsi que les exigences technologiques nécessaires pour collecter ces mesures.

L'OBJECTIF EST DE METTRE EN PLACE DES CONTRÔLES PROACTIFS ET AINSI D'ÉLIMINER L'INTRODUCTION DE DONNÉES DÉFAILLANTES DANS L'ENVIRONNEMENT

- Préparer le rapport d'évaluation de la QD – Le processus de documentation dans un rapport unique de la corrélation entre les incidences opérationnelles et les anomalies informationnelles au moyen de différentes méthodes de mesure offre un « point fixe » aux consommateurs de données opérationnelles pour ce qui est de l'état actuel de la qualité des données et propose une ligne de départ pour la détermination des niveaux d'amélioration cibles.

Mesures et métrologie de la qualité des données

Après l'évaluation destinée à identifier les secteurs pouvant bénéficier d'une amélioration des données, l'étape suivante consiste à synthétiser les résultats de l'évaluation pour restreindre le cadre en se concentrant sur les éléments de données considérés comme essentiels suivant les besoins des utilisateurs fonctionnels. La définition de la métrologie des performances au moyen d'une carte de notation de la qualité des données exige la présence de processus visant à la détermination des dimensions, unités de mesure correspondantes et seuils d'acceptabilité ainsi qu'à la présentation de paramètres quantifiables portant sur les consommateurs de données opérationnelles.

Pour reprendre notre exemple, une fois qu'il a été déterminé que les problèmes d'adresse ont une incidence négative sur la livraison optimale des articles expédiés, nous pouvons cibler les mesures de la qualité de données sur les paramètres associés aux éléments critiques qui contribuent aux échecs de livraison. Certains articles risquent de ne pas pouvoir être livrés parce qu'il manque le nom de la rue ou encore parce que le code postal est incorrect. Dans le premier cas, il s'agit d'un manque d'informations alors que dans le deuxième, il s'agit plutôt d'un problème de cohérence des données de référence définies. Les mesures associées aux dimensions d'intégralité et de cohérence des données peuvent être définies au moyen des règles de validation de chaque adresse. Les mesures en résultant peuvent être présentées comme des paramètres aux utilisateurs opérationnels du service de traitement pour estimer le rapport entre les adresses non valables et la hausse des coûts.

Les différents aspects de cette pratique sont les suivants :

- Sélectionner les dimensions de la qualité des données – Une dimension décrit un contexte et un cadre de référence ainsi que différentes suggestions en termes d'unités de mesure. Les dimensions couramment mesurées sont l'intégralité, la cohérence, l'à-propos et l'individualité des données, mais l'éventail de dimensions possibles n'est limité que par la capacité de fournir une méthode de mesure. Au cours de ce processus, les analystes de la qualité des données sélectionnent les dimensions à mesurer et considèrent les outils, techniques et compétences nécessaires à la capture des mesures. Il résulte de ce processus un ensemble de mesures spécifiques pouvant être combinées pour aboutir à une métrologie qualitative.
- Définir la métrologie qualitative – Une fois les dimensions applicables aux consommateurs de données opérationnelles et les mesures spécifiques identifiées, l'analyste est en mesure de créer des paramètres communicables précis pouvant être présentés aux gestionnaires de données. Il peut s'agir de paramètres élémentaires composés de règles mesurées directement ou de paramètres plus complexes reprenant les moyennes pondérées des notes rassemblées. D'autres aspects sont les schémas et les méthodes permettant une progression descendante dans les données imparfaites pour réaliser une analyse des causes fondamentales.
- Définir les règles de validité des données – Le processus d'évaluation met en lumière les anomalies potentielles qui sont examinées par les utilisateurs opérationnels pour identifier les mesures de qualité des données qui donne lieu aux paramètres de qualité des données. Toutefois, pour s'éloigner de l'approche réactive visant à corriger les erreurs dès qu'elles se manifestent au niveau de l'interface des utilisateurs finaux, l'entreprise doit incorporer les contrôles de données dans le processus de développement de l'application de sorte que les erreurs puissent être identifiées et rectifiées dès qu'elles se produisent. Au cours de ce processus, les analystes mettent en place des règles de validité des données qui sont ensuite intégrées aux applications de gestion sous forme de contrôles pour vérifier que les données sont bien conformes aux attentes, de bout en bout du flux d'informations.

Cinq pratiques fondamentales en matière de qualité des données

- Définir des seuils d'acceptabilité – Une fois les dimensions et paramètres validés, les utilisateurs opérationnels sont consultés pour faire part de leurs seuils d'acceptabilité. Lorsqu'une note est en deçà du seuil d'acceptabilité, cela signifie que les données ne respectent pas les attentes. L'intégration de ces seuils aux méthodes de mesure complète l'élaboration de la métrologie de la qualité des données.
- Élaborer une carte de notation de la qualité des données – Une carte de notation présente les notes aux gestionnaires qui observent les jeux de données opérationnelles. Les notes peuvent être capturées dans un référentiel sur une longue période pour permettre l'analyse des tendances et mettre en lumière les améliorations ou l'absence d'amélioration. Le processus d'élaboration des cartes englobe la définition des paramètres, les processus de mesure, la pondération, la méthode de capture et de stockage des notes ainsi que la composition des outils et des technologies en vue de la livraison et de la présentation.

Qualité des données et cycle de développement du système

Trop souvent, la qualité des données vient après coup. Le personnel corrige les erreurs à mesure qu'elles sont constatées au lieu d'éliminer ce qui cause l'imperfection des données. Du fait que la qualité ne peut pas être simplement ajoutée après coup, une fois que les processus d'identification de l'incidence de la mauvaise qualité des données sur les opérations ainsi que la capacité à définir des règles d'inspection et de surveillance sont en place, l'étape suivante consiste à incorporer cette inspection directement dans les applications de gestion. Essentiellement, la pratique suivante consiste à établir les moyens par lesquels la gestion de la qualité des données est pensée et conçue au travers de l'architecture de l'entreprise.

Toutefois, du fait que les approches traditionnelles en matière d'analyse et de conception des besoins du système visent à se concentrer sur les besoins fonctionnels des applications transactionnelles et opérationnelles, les besoins en informations des processus en aval sont

largement ignorés et ne font surface que bien après la mise en production des applications. La conception de la gestion de la qualité des données dans l'entreprise appelle à une reformulation de l'analyse des besoins, avec un nouvel axe sur les besoins en informations horizontales et en aval, au lieu de ne prendre en compte que les besoins fonctionnels immédiats.

Pour reprendre notre exemple dans lequel les adresses non valables entraînent une hausse des coûts d'expédition, deux approches sont possibles. L'approche réactive consiste à soumettre toutes les adresses à un processus de nettoyage et de perfectionnement avant de générer une étiquette d'expédition. Cette démarche, si elle est susceptible de réduire certains des coûts, ne garantit pas l'exactitude de tous les enregistrements. En revanche, si les règles de validité sont connues, elles peuvent être intégrées directement dans l'application à la création des adresses. En d'autres termes, la validation et la correction des adresses au moment où les clients les saisissent évitent tout simplement l'introduction d'erreurs dans l'environnement !

Les processus qui contribuent à cette pratique sont les suivants :

- Analyse des besoins en terme de qualité des données – Au cours de ce processus, les analystes synthétisent les attentes en termes de qualité pour les jeux de données consommés sur la base de l'analyse de l'incidence opérationnelle, de la détermination des dimensions de la qualité des données et des aspects de hiérarchisation se rapportant à la faisabilité aussi bien qu'aux répercussions sur le système. Pour chaque application de gestion, le flux d'information est retracé jusqu'à la création ou l'acquisition des données. La représentation ainsi obtenue fait l'objet d'une enquête visant à déterminer les points les plus propices où insérer les formalités d'inspection des données. L'analyste peut proposer la mise en place de contrôles aux endroits où les données sont extraites, transformées ou échangées. L'analyste peut proposer la mise en place de contrôles aux endroits où les données sont extraites, transformées ou échangées. Ces contrôles déclencheront des événements de notification lorsqu'une infraction est constatée en aval.

LA NORMALISATION DE LA FAÇON DONT LA QUALITÉ DES DONNÉES EST DÉPLOYÉE ET L'UTILISATION D'OUTILS ADAPTÉS ASSURENT LA FIABILITÉ ET LA VALEUR DES INFORMATIONS.

- Perfectionnement du CDS pour la QD – Incorporation d'une validation, d'une inspection et d'une communication des données dans les processus de gestion et dans l'application de gestion correspondante en ajustant le Cycle de Développement du Système (CDS) de façon à ce que les besoins en données organisationnelles puissent être sollicités et intégrés dans la phase d'analyse des besoins du développement du système. Ce processus s'axe sur l'appartenance des données et la façon dont le modelage des processus de gestion peut servir à élaborer les besoins en informations et les besoins fonctionnels. Puisque les utilisateurs en aval, tels que les consommateurs de rapports sur l'informatique décisionnelle, dépendent des données collectées au cours des activités opérationnelles, il est nécessaire de collecter formellement les besoins en données au cours du processus CDS de façon à ce que les besoins en données organisationnelles puissent être sollicités et intégrés dans la phase d'analyse des besoins du développement du système.
- Intégrer les méthodes d'amélioration de la qualité des données – La capture des besoins de l'entreprise au cours des phases d'analyse des exigences et de conception du cycle de développement habilite l'équipe de développement à incorporer directement la qualité et la correction des données dans l'application. C'est-à-dire qu'il devient possible de valider les valeurs et les enregistrements à la saisie dans l'environnement (par acquisition ou par création) ou lors des transferts entre phases de traitement, de vérifier l'acceptabilité et, soit de renvoyer les données au fournisseur pour résolution, soit d'appliquer les ajustements ou corrections à la volée.

Amélioration de la qualité des données opérationnelles

Une fois les besoins rassemblés, les règles de validation définies et les méthodes de mesure de la performance stipulées, l'étape suivante consiste à établir le contrat entre les fournisseurs de données et les consommateurs de données portant sur le niveau de service nécessaire au maintien de données de haute qualité.

Dans notre exemple, les adresses sont validées par rapport à un ensemble de normes définies, soit celles utilisées spécifiquement par les agences postales des différents pays,

soit les normes « de facto » employées par les professionnels pour assurer la distribution. Ces normes définissent des tables de référence, ou autres artefacts de métadonnées, veillant à la validité d'un lieu de livraison.

En combinant les règles de validité des données et les métadonnées consignées, les analystes sont en mesure de documenter le niveau d'acceptabilité des données de type adresse attendues par les utilisateurs opérationnels. La performance des activités rectificatives peut alors être mesurée pour garantir que la qualité des données est acceptable.

La pratique consistant à établir une convention sur les niveaux de service en termes de qualité englobe les tâches suivantes :

- Gestion de la normalisation de données – L'absence de cadre de référence commun, de définitions des termes courants et d'un format convenu d'échange ne facilite pas la compréhension entre les différentes parties. Cela est plus particulièrement vrai lorsque des informations spécifiques doivent être communiquées au niveau de plusieurs applications de gestion. Cette difficulté suggère la nécessité de mettre en place une normalisation du partage des données. Une norme de données est une convention entre des collaborateurs sur la définition de termes de gestion courants, leur dénomination et leur représentation au niveau des données ainsi qu'un ensemble de règles décrivant le stockage, l'échange, le formatage ou la présentation des données. Ce processus décrit les politiques et les procédures de définition des règles et de conclusion d'accords sur les éléments de données standard.
- Gestion active des métadonnées – Du fait que l'utilisation des éléments de données et leurs concepts sous-jacents influencent la façon dont les applications de gestion s'exécutent, le référentiel de métadonnées d'une entreprise peut jouer le rôle de « centre de commande » pour maîtriser et gérer la façon dont ces applications se servent des concepts de données courants. Outre le besoin de collecter des détails techniques standard portant sur les différents éléments de données potentiellement disponibles, un référentiel de métadonnées est utile pour
 - > Déterminer les utilisations opérationnelles de chaque élément de données

Cinq pratiques fondamentales en matière de qualité des données

- > Déterminer la définition des éléments de données qui se rapportent à des concepts similaires
- > Identifier les applications se rapportant à ces concepts de données,
- > Examiner comment chaque élément de données et concepts connexes sont créés, lus, modifiés ou supprimés par les différentes applications
- > Documenter les caractéristiques de la qualité des données et noter les emplacements d'inspection et de surveillance au sein du flux des processus de gestion
- > Récapituler comment toutes ces utilisations sont connectées

Par conséquent, un programme de gestion des métadonnées opérationnelles permettant de mettre en place dans toute l'entreprise le niveau de normalisation souhaité est un composant indispensable d'une architecture d'un système d'information.

- Inspection et surveillance de la qualité des données – La disponibilité de règles de validation se trouve à la base de l'inspection et de la surveillance de la qualité des données. L'insertion de sondes d'inspection et la surveillance de la qualité des données avancent les moyens d'identification des irrégularités et de notification aux personnes responsables de façon à mettre en œuvre les mesures correctives convenues. Les mécanismes d'inspection et de surveillance des données ainsi que les flux de travaux correspondants doivent être définis à des fins d'inspection ainsi que pour garantir que les éléments de données, les enregistrements et les jeux de données respectent bien les besoins en aval.

Ce processus implique la définition de routines d'inspection de la qualité pouvant comprendre des processus automatisés aussi bien que manuels. Les processus automatisés peuvent englober le résultat des contrôles de validation exécutés lors du traitement des applications, du profilage des données, de l'automatisation des analyses des données, du traitement personnalisé ou encore avec les outils ETL. L'inspection manuelle peut faire intervenir l'exécution de requêtes ou de rapports sur les sources de données ou encore l'obtention d'échantillons de données qui sont ensuite sujets à étude. Les procédures d'inspection

sont définies pour chaque dimension de qualité des données pertinentes. Ces méthodes d'inspection sont personnalisées pour chaque système, selon les besoins.

- Définir les conventions sur les niveaux de services en matière de qualité des données - Une convention sur les niveaux de service est un contrat entre un prestataire de service et ses clients qui précise les responsabilités du prestataire par rapport aux différents aspects mesurables de ce qui est fourni, tels que la disponibilité, le rendement, le temps de réponse en cas de problème, etc. Une convention sur les niveaux de service en matière de qualité des données, ou CNS QD, est un accord qui précise les attentes du consommateur de données en termes de règles de validité et niveaux d'acceptabilité ainsi que les attentes raisonnables en termes de réponse et de mesures correctives lorsque des erreurs et des défaillances de processus sont découvertes. Les CNS QD peuvent être établies dans toutes les situations où un fournisseur de données procure des données à un consommateur de données.

Ce processus précise les attentes dans le domaine des aspects mesurables portant sur une ou plusieurs dimensions de qualité des données (notamment, l'exactitude, l'intégralité, la cohérence, l'à-propos), en suivant les consignes des processus déjà présentés. La convention sur les niveaux de service précise également ce que signifie la conformité à ces attentes et décrit le flux de travaux à exécuter lorsque ces attentes ne sont pas remplies. Les problèmes signalés sont hiérarchisés et les personnes appropriées sont notifiées afin de prendre les mesures nécessaires à la résolution avant que les opérations ne soient négativement affectées.

Suivi des problèmes, mesure corrective et amélioration

L'opérationnalisation de la convention sur les niveaux de services implique que des processus de signalisation, de consignation et de suivi des problèmes de qualité des données émergents ou constatés soient en place. Des structures de rapport des incidents peuvent être adaptées à cet effet pour permettre aux gestionnaires de données de se concentrer sur l'évaluation des causes fondamentales des problèmes de données et de proposer un plan de rectification, pouvant aller de la refonte des processus à la simple correction des données.

EN COMBINANT DES PRATIQUES SAINES DE GESTION DES DONNÉES AVEC UNE PLATEFORME TECHNOLOGIQUE ADAPTÉE, UNE ENTREPRISE EST EN MESURE DE COMPLÈTEMENT INCORPORER LA QUALITÉ DES DONNÉES DANS SON ARCHITECTURE

Le suivi, la consignation et la gestion des problèmes garantissent qu'aucun incident constaté ne soit laissé pour compte. Dans notre exemple, à chaque retour du fait d'un problème de qualité, une analyste étudiera l'erreur pour déterminer la source du problème, considérera si elle est le résultat de l'omission d'une étape de validation ou déterminera s'il y a une nouvelle cause fondamentale susceptible d'entraîner la création de nouvelles règles de validation à intégrer au flux de processus opérationnel.

Cette pratique englobe les tâches suivantes :

- Rapport et suivi des problèmes de qualité des données – La mise en œuvre d'une convention sur les niveaux de service fait appel aux processus de rapport et de repérage des problèmes de qualité des données ainsi qu'à toutes les activités de suivi. "Un système de consignation et de repérage des problèmes favorise une évaluation plus formelle ainsi que l'établissement d'un diagnostic initial de « mauvaises données. D'autre part, la présence d'un système de repérage permet au personnel d'être plus efficace dans l'identification et, par extension, dans la rectification des problèmes associés aux données. Le repérage des incidents alimente également l'établissement de rapports de rendement mesurant notamment le temps moyen nécessaire à la résolution des problèmes, la fréquence des problèmes, le type de problèmes rencontrés, la source des problèmes et les démarches prises pour corriger ou éliminer les problèmes.
- Analyse des causes fondamentales – Intégrées aux applications de gestion en tant que mesures de contrôle, les règles de validation des données peuvent déclencher des notifications lorsqu'une erreur survient. À ce stade, il incombe aux gestionnaires de non seulement corriger les données, mais aussi d'identifier la source d'introduction des erreurs dans les données. Le processus d'analyse des causes fondamentales se sert d'outils et de techniques d'inspection et de surveillance pour isoler la phase de traitement dans laquelle l'erreur s'est produite et pour étudier les processus de gestion afin de déterminer la cause fondamentale ultime des erreurs.
- Nettoyage des données – La rectification des erreurs dans les données est instinctivement réactive. Elle incorpore des processus permettant de corriger les erreurs afin de respecter les limites d'acceptabilité,

plus particulièrement lorsque la cause fondamentale est indéterminable ou si les modifications de processus se trouvent en dehors du domaine administratif de l'intendant. Les corrections doivent faire l'objet d'une socialisation et d'une synchronisation avec tous les consommateurs et les fournisseurs de données, plus particulièrement si les données sont utilisées dans des contextes de gestion différents. Par exemple, un accord sur les modifications doit être en place lors de la comparaison des données communiquées et des résultats composites pour éviter que des chiffres divergents, ne disposant pas d'explications, ne fassent l'objet d'un rapprochement, ce qui serait une perte de temps.

- Correction des processus – Malgré l'existence d'un mécanisme administré de correction des mauvaises données, le fait que des erreurs se produisent suggère des anomalies au niveau des processus qui doivent être examinés et éventuellement rectifiés. La correction englobe un processus administré d'évaluation du flux de production de l'information, du déroulement des processus opérationnels ainsi que la détermination de ce qui doit être fait pour les améliorer afin de réduire ou d'éliminer l'introduction d'erreurs.

Pratiques en matière de qualité des données et services de données essentiels

La mise en place d'un programme de gestion des données de la qualité va au-delà de l'achat d'outils de nettoyage ou l'établissement d'un comité de gouvernance des données. En effet, la mise en place d'un tel programme ne se limite pas à la consignation d'une série de processus. Un cycle d'évaluation itérative, la planification, l'exécution et la gestion du rendement dans le domaine de la qualité des données exigent la présence de processus reproductibles qui allient compétences humaines et outils appropriés. En outre, le personnel choisi pour participer au programme doit disposer des outils adéquats pour pouvoir passer de la théorie à la pratique. Cela suggère une combinaison de technologie et de formation à cette technologie, au travers de services de données tels que :

- L'intégration des données, pour mettre en place des moyens de récupération et de transformation des données entre différents systèmes.

Cinq pratiques fondamentales en matière de qualité des données

- Le profilage des données, utilisé dans l'évaluation de la qualité des données, la validation, l'inspection et la surveillance des données.
- L'analyse syntaxique, la normalisation et la résolution d'identité, utilisées dans la validation des données, l'identification des erreurs la normalisation et la correction des données.
- Le couplage et la fusion des enregistrements, utilisés également dans l'identification des erreurs et la résolution des variances et des erreurs ultérieures.

Un sous-ensemble de services de données essentiels existe pour la normalisation des pratiques saines de gestion des données. La normalisation de la façon dont la qualité des données est déployée et l'utilisation d'outils adaptés assurent la fiabilité des informations. Lors du développement ou de la refonte de l'architecture d'une entreprise, la mise en œuvre des pratiques de qualité fondamentales permet de réduire la complexité du cadre de gestion des données ; les efforts et les risques sont réduits et le degré de confiance dans les informations de l'entreprise est élevé.

PITNEY BOWES BUSINESS INSIGHT : RÉPONSES ET SOLUTIONS

The Pitney Bowes Spectrum™ Data Integration Solution	The Pitney Bowes Spectrum™ Enterprise Data Quality Solution	The Pitney Bowes Spectrum™ Enterprise Location Intelligence Solution	The Pitney Bowes Spectrum™ Business Services
<p>Data Services for Oracle</p> <p>Data Services for Siebel</p> <p>Data Services for SFDC</p> <p>Data Services for SAP</p> <p>Sagent Data Flow</p>	<p>Profiler Plus</p> <p>Monitor Plus</p> <p>Universal Addressing Module</p> <p>UAM International Enhanced</p> <p>Address Now Module</p> <p>Data Normalization Module</p> <p>Universal Name Module</p> <p>Advanced Matching Module</p>	<p>Enterprise Geocoding Module</p> <p>Location Intelligence Module</p> <p>Enterprise Routing</p>	<p>Enterprise Tax</p> <p>GeoConfidence</p> <p>Data Quality Connector for mySAP</p> <p>Data Quality Connector for Siebel</p> <p>Global Sentry</p>
<p align="center">The Pitney Bowes Spectrum™ Technology Platform</p> <p align="center">ENTERPRISE DESIGNER • INTERACTIVE DRIVER • MANAGEMENT CONSOLE • CLIENT APIs • WEB SERVICES</p>			

ÉTATS-UNIS

One Global View
Troy, NY 12180
Tél. : 518.285.6000
800.327.8627
www.pbinsight.com
pbbi.sales@pb.com

France

Immeuble Le Dièse
134, rue Danton
92593 Levallois-Perret Cedex
+33 1 46 17 54 00
www.pbinsight.fr
info.pbbifrance@pb.com

EUROPE/ROYAUME-UNI

Minton Place, Victoria Street
Windsor, Berkshire SL4 1EG
Royaume-Uni
+44.1753.848.200
www.pbinsight.co.uk
pbbi.europe@pb.com

ASIE PACIFIQUE/AUSTRALIE

Level 7, 1 Elizabeth Plaza
North Sydney NSW 2060
Australie
+61.2.9437.6255
pbbi.australia@pb.com
pbbi.sea@pb.com
pbbi.china@pb.com