

# Cinq étapes-clés pour des données d'entreprise performantes

*La qualité et la gestion des données reposent autant sur une organisation et des processus dédiés que sur la technologie.*

## Introduction

Pour qu'une décision soit opportune au niveau de l'entreprise, elle doit s'appuyer sur des données correctes. Qu'il s'agisse de décisions opérationnelles ou encore stratégiques, chaque acteur, depuis la Direction générale jusqu'au point de vente, a besoin de disposer d'informations fiables et exactes pour prendre les décisions appropriées - connaître son parc client pour le cibler en regard d'un nouveau produit, statuer sur l'opportunité d'une mise sur le marché en regard de la typologie de mes clients, etc.. Toutefois, il convient de reconnaître que de mauvaises données conduisent inéluctablement à de mauvaises décisions, à des stratégies inadéquates ainsi qu'à un service client médiocre.

Attendu que les entreprises mettent en œuvre des initiatives basées sur les données, telles que les ERP (Progiciels de Gestion Intégrés), la gestion de la relation client (CRM) ou autres, il est évident que les problématiques de qualité des données existent dans toute entreprise. Les appréhender efficacement - et instaurer des contrôles pour maintenir cette qualité - est essentiel au bon fonctionnement de chacune de ces applications.

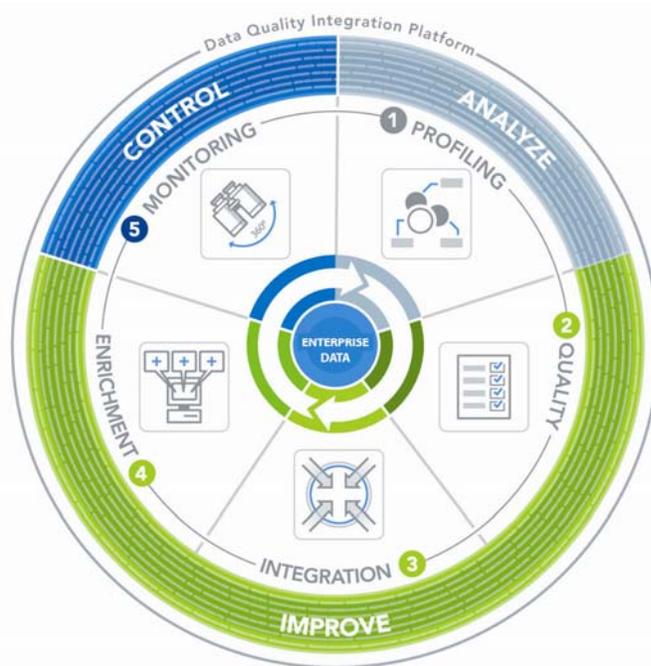
Il existe bon nombre de raisons pour que les données soient pauvres ou inutilisables, que ce soit du fait d'une simple erreur humaine ou de l'absence de standards à travers les systèmes ou les organisations. Pour pallier les problèmes liés à des données impropres, les organisations doivent construire un socle technologique afin de les identifier et les éliminer.

Cependant, la seule technologie ne peut résoudre ces problèmes. Il est donc impératif de se doter des compétences et des processus idoines. Ce livre blanc a pour objet de démontrer comment on peut fédérer efficacement les hommes, les processus et la technologie au travers d'une méthodologie en cinq étapes qui permet aux organisations d'analyser, améliorer et contrôler les données d'entreprise.

## Les cinq étapes-clés pour des données d'entreprise performantes

*Les cinq composants-clés d'un programme de gestion des données : profilage, mise en qualité, intégration, enrichissement et monitoring.*

L'élaboration d'un système d'information « corporate » performant suppose davantage que la seule technologie. S'il est crucial de pouvoir disposer d'une technologie capable de couvrir un large spectre en termes de qualité de donnée, il s'avère également essentiel de s'appuyer sur une méthodologie adéquate pour pouvoir la gérer efficacement. DataFlux met en œuvre une méthodologie articulée autour de cinq étapes, dédiée à la qualité et la gouvernance des données d'entreprise, qui est le fruit d'une expérience de plus de dix ans et de milliers d'implémentations. Ces cinq étapes-clés sont le profilage (data profiling), la mise en qualité (data quality), l'intégration (data integration), l'enrichissement (data enrichment) et le monitoring (data monitoring). Ces cinq étapes offrent une approche pratique et éprouvée qui permet aux organisations d'analyser, d'améliorer et de contrôler leurs données ainsi qu'un cadre efficace à l'ensemble des processus d'amélioration des données. Lorsqu'elle est appliquée, cette méthodologie permet d'avoir une vue unifiée de toutes les typologies de données, dont les données client, produit ainsi que, par exemple, les informations fournisseurs à tous les niveaux de l'entreprise.



### Le profilage

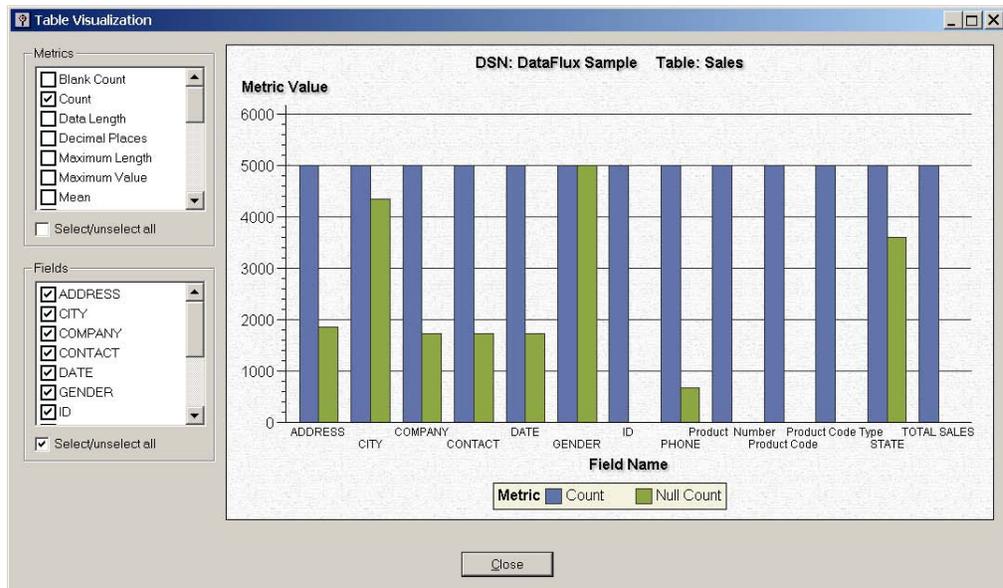
Première étape de la méthodologie, le profilage (data profiling) consiste en l'examen de la structure, des relations et des contenus des données existantes afin de déterminer leur état. Ce procédé concourt à l'élaboration de meilleures solutions pour la correction ou l'harmonisation des informations. Au cours de l'étape de profilage, les données sont examinées sous trois angles différents, conduisant à une vision claire de la nature et de la teneur des problèmes.

Le profilage constitue une étape essentielle dans l'analyse des problématiques de qualité des données au travers de la découverte:

- de leur structure - Les modèles de données correspondent-ils aux modèles attendus ? Les données correspondent-elles aux métadonnées définies ?
- de leur contenu - Les valeurs des données sont-elles complètes, précises et sans ambiguïté ? La donnée est-elle en conformité avec les règles en vigueur ?
- de leurs relations - Existe-t-il une corrélation entre les enregistrements d'une table ou de plusieurs tables ? Existe-t-il des données redondantes ?

La Figure 2 montre un rapport type en matière de découverte de l'information. Cette analyse fait ressortir les éléments invalides ou caduques des différents champs au sein de la base de données. A noter que le total de 5 000 dossiers demeure cohérent. Toutefois, certains champs montrent un nombre élevé d'éléments non valides. Sachant que le pourcentage d'éléments manquants dans plusieurs champs-clés est important - adresse, ville, état et genre, en particulier - les analystes peuvent être immédiatement alertés quant à la nature et l'étendue des problèmes de qualité des données.

*Le profilage vise à déterminer l'état de la qualité des données d'entreprise.*



Le profilage permet d'obtenir des informations quantitatives détaillant les forces et les faiblesses des données d'entreprise... Une connaissance particulière qui peut, par conséquent, servir de point de départ aux processus d'amélioration de ces dernières. Les informations obtenues à partir du profilage servent de fondation à la phase suivante : la mise en qualité (data quality).

*La phase de « mise en qualité » tire parti des enseignements issus du profilage pour appréhender les problématiques de données et en améliorer la qualité.*

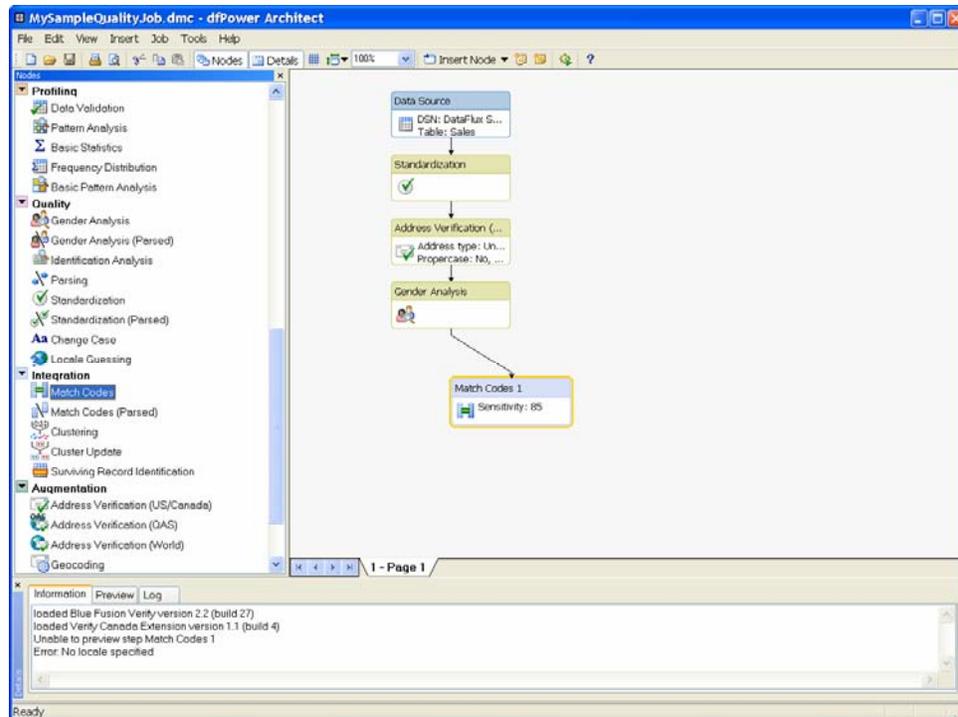
## La Mise en qualité

La phase de mise en qualité met en pratique les enseignements obtenus au cours du profilage pour créer les processus d'amélioration des données. Cette phase contribue à corriger les erreurs, standardiser les informations et valider les données en regard des règles de l'entreprise. Pour ce faire, on met en œuvre différentes mesures dont les plus communément adoptées incluent :

- La standardisation des données : pour prendre en charge et corriger les multiples permutations des données. Par exemple, ACME Manufacturing Corporation peut coexister avec Acme Mftg Corp, ACME et ACME Manufacturing. Un rapprochement intelligent permet une standardisation de ces valeurs pour créer une nomenclature unique.
- La standardisation des formats : pour créer des formats standards quelque soit le lieu de stockage de la donnée. Certains attributs, comme les numéros de téléphone ou de Sécurité Sociale, suivent des formats facilement modélisables. D'autres, comme les données produit ou les données thématiques, peuvent suivre différents standards suivant les industries ou les entreprises. La standardisation des formats permet d'homogénéiser la représentation de ces données.
- La vérification d'adresse : pour confirmer la validité des adresses. Par exemple, la solution DataFlux peut déterminer si la ville et le département sont cohérents avec le code postal. Si ce n'est pas le cas, il sera modifié en s'appuyant sur les données postales de référence pour créer une adresse exacte.

La Figure 3 montre comment la technologie peut contribuer à la création de processus métiers qui traiteront les problèmes de qualité étape par étape. Dans cet exemple, le workflow lit la table source, standardise certaines informations, vérifie l'adresse, etc.. L'utilisateur prend ainsi des mesures pour résoudre chacun des problèmes identifiés lors du profilage.

*Il est essentiel de rechercher et de traiter les causes de la non-qualité*



L'approche choisie varie en fonction de chaque attribut, et relève de l'arbitrage des acteurs métiers.

Par ailleurs, il est important de souligner que la mise en qualité améliore les données, mais ne traite pas les causes. Si l'entreprise s'intéresse réellement à l'amélioration de la qualité, elle doit également rechercher les raisons pour lesquelles les données sont incorrectes et mettre en œuvre le plan d'action qui s'impose, avec des impacts tant organisationnels que liés aux process. Une fois la mise en qualité terminée, l'entreprise a une vision claire des actions à prendre pour amener la complétude et l'exactitude de l'ensemble des données à un niveau acceptable. La prochaine étape concerne l'intégration de sources multiples.

## L'intégration

Quelque soit le type de données, on retrouve les mêmes objets et les mêmes attributs dans de multiples bases de données. L'intégration s'avère utile lorsqu'on tente de rationaliser les données issues de différentes sources. L'intégration permet de rassembler ces informations pour les uniformiser au sein d'un référentiel unique. Prenons l'exemple d'une société qui possède deux fichiers produit : l'un provient du système ERP basé aux Etats-Unis, l'autre est issu de la base de données européenne. La société vend les mêmes produits dans les deux régions, mais les produits portent différents noms suivant les bases de données. De plus, le produit, la marque et les caractéristiques de chaque fiche

*La technologie de rapprochement intelligent contribue largement à la simplification de l'intégration*

*L'intégration agrège les informations issues de sources multiples en un référentiel unique*

*Enrichir les données d'entreprise avec des données tierces est source de valeur.*

ont été saisis par différentes personnes. Le premier défi inhérent à l'intégration est de constater l'existence du même produit au sein de deux bases de données différentes et le second défi consiste à consolider les données pour obtenir une vue unifiée du produit.

Les données clients, quant à elles, comportent un champ commun - comme le numéro de contribuable, le NIR ou encore un DUNS number - qui peut aider à identifier la domiciliation. Lorsque cela arrive, les entreprises peuvent rapidement et facilement identifier plusieurs informations d'un même client.

Cela n'est souvent pas le cas avec les données produit. Dans l'exemple précédent, la technologie d'intégration a permis à l'entreprise de faire le lien entre les deux systèmes et consolider les données. Le rapprochement intelligent a identifié une caractéristique produit commune aux deux fiches. La fiche américaine contenait les caractéristiques, le nom de marque et l'identifiant produit dans un seul champ et plusieurs formats de donnée. La fiche européenne contenait seulement les caractéristiques produit, avec différents formats et abréviations également. Afin de faire le lien entre elles, l'intégration a été utilisée pour :

- Faire l'analyse des caractéristiques en s'appuyant sur les spécificités du produit et le nom de marque à partir de la fiche américaine.
- Faire le rapprochement entre les noms de marques.
- Faire la comparaison entre les différentes spécificités du produit (formes courtes, abréviations, etc.).
- Etablir les correspondances phonétiques entre les données recueillies.
- Afficher les rapports d'adéquation produit.

A l'issue de l'intégration, l'entreprise possède une vision claire de l'ensemble des données relatives à un sujet. Les mêmes principes et techniques peuvent être appliqués à tout type de données pour la création de référentiels unifiés. L'étape suivante consiste à augmenter la valeur des données en les enrichissant d'informations supplémentaires.

## L'enrichissement

L'enrichissement consiste en l'intégration de données externes afin de donner davantage de valeur aux données existantes. Pour ce faire, plusieurs moyens existent.

L'enrichissement peut permettre d'améliorer la connaissance du consommateur, de connaître quel est son pouvoir d'achat et de mieux le fidéliser. Autre exemple, le géocodage avancé peut permettre de déterminer l'ensemble des maisons d'un quartier, construites après 1980 et d'une certaine valeur immobilière, et fournir ainsi des

*L'enrichissement peut être utilisé pour classifier l'information, au travers de standards tels que eCI@ss ou UNSPSC.*

informations pour cibler les offres sur certains produits. Grâce à l'analyse de certaines caractéristiques, on peut également obtenir une meilleure compréhension du comportement des clients. Intégrer ces données aux données déjà connues permet de segmenter les consommateurs plus efficacement et d'identifier ainsi des opportunités qualifiées.

La Figure 4 illustre la manière dont on peut améliorer une adresse en comparant les données initiales avec un référentiel postal externe.

Submitted Data	Standardized Data	
DataFlux	DataFlux Corporation	
940 Cary Pkwy	940 Cary Pkwy, <b>Ste. 201</b>	
Cary, NC	Cary, NC 27513-2792	
New Data Points Added		
	Carrier Route ID	34
	Delivery Point ID	1
	County Number	183
	County Name	Wake
	Congressional District	4
	Latitude	35.8306
	Longitude	-78.7841
	Census Block Group	371830535122

L'enrichissement couvre bien entendu bien d'autres aspects que l'adressage postal. De nombreux types de données, comme les données produit, d'inventaire ou financières, peuvent être enrichies au travers de sources externes. On peut ainsi classifier automatiquement des produits en s'appuyant sur une nomenclature standard (UNSPSC, eCI@ss) pour normaliser les codes produits, et simplifier l'achat et la vente de marchandises. Le fait d'intégrer cette nomenclature peut permettre de garantir l'exactitude de votre inventaire, améliorer l'efficacité de vos processus de commande et ce dans un contexte international. L'enrichissement contribue à la gestion des problématiques de conformité en vérifiant que les listings client ou les transactions répondent aux exigences des réglementations gouvernementales ou industrielles... Un moyen de lutte notamment contre le blanchiment d'argent, pour la détection des fraudes ou l'intégration des fichiers d'appels non sollicités.

## Le Monitoring

Le Monitoring s'avère essentiel dans la mise en œuvre d'un programme de gestion des données en ceci qu'il veille à ce que l'amélioration et l'enrichissement des données ne soient pas affectés par des retours d'erreurs latentes. Un monitoring actif permet aux entreprises d'isoler et corriger les causes des problèmes de qualité.

Le monitoring peut prendre nombre de formes différentes. En la matière, la version la plus basique automatise un rapport de profilage - une analyse périodique des données qui recherche les éléments exceptionnels et les données non-standards. Déclencher des événements sur seuil peut générer un e-mail ou un système d'alerte auprès du personnel dédié ou du département en charge des données. Une autre méthode de monitoring, plus dynamique et plus efficace, consiste à faire appliquer les règles métier en s'appuyant sur des web services. Ainsi, les mêmes règles utilisées pour nettoyer et améliorer les données en différé sont appliquées aux données lors de leur intégration au Système d'Information. Les avantages de cette démarche sont de donner la capacité à :

- Faire appliquer les règles de gouvernance - S'assurer que les règles métier spécifiques sont suivies.
- Comprendre et affiner les process stratégiques - Les violations ou exception faites aux règles métier sont stockées dans un référentiel à des fins de traçabilité et d'analyse.
- Invoquer des événements pour corriger les données - Le système de monitoring opérationnel peut déclencher des événements pour corriger automatiquement les données, envoyer des alertes par e-mail aux utilisateurs, sauvegarder les données dans un référentiel ou dans une simple table.

## En résumé

La qualité de toute décision métier dépend de celle des données. En l'absence de données riches, exactes et fiables transversales à l'entreprise, une organisation peut rapidement se fourvoyer en conclusions trompeuses, erronées et potentiellement dangereuses. La voie de la réussite en matière d'analyse des données commence par la mise en œuvre de la méthodologie qui supporte chacun de ces éléments constitutifs - le profilage, la qualification, l'intégration, l'enrichissement et le monitoring. La technologie innovante alliant qualité et intégration de la donnée permet de supporter les initiatives des sociétés en terme de Système d'Information à un niveau sans précédent, en produisant une vision unifiée des données client, produit, fournisseur et autres actifs. Grâce à la conjugaison d'une méthodologie et d'une technologie adéquate, une organisation peut mettre en place une démarche qualité permanente dans l'entreprise. Les décisions sont plus intelligemment prises et l'avantage compétitif d'autant amélioré.