

Data quality: garantir la cohérence et la fiabilité des données pour optimiser les performances

Maîtrisez vos données et améliorez vos prises de décisions: découvrez comment un catalogage précis et une stratégie de qualité des données validée peuvent transformer votre gestion de l'information.

Stéphane LE LIONNAIS

Expert en gouvernance de données
et co-fondateur de Dawizz



Un contexte où la gouvernance des données est de plus en plus complexe et nécessite une meilleure maîtrise

La véritable valeur d'une organisation, c'est son patrimoine data. **Cependant, saviez-vous que les informations des bases de données proviennent de données structurées, qui ne représentent que 20% de l'ensemble des données d'une organisation?** Saviez-vous également qu'un simple e-mail contenant du texte et une image est considéré comme un fichier à données non structurées? **Il est de fait inutilisable (en l'état) par l'Intelligence Artificielle ou la Business Intelligence, par exemple.** C'est regrettable, car leur utilisation devient un avantage concurrentiel indéniable, mais ils nécessitent une grande quantité de données pour être efficaces.

C'est un paradoxe, car en 2025, le monde comptera environ 175 milliards de téraoctets de données (ou 175 000 milliards de Go), mais seulement 0,5% de ces données sont actuellement utilisées. De plus, il faut tenir compte du fait que ces 0,5% comprennent des doublons, des données non-conformes, non sécurisées, non documentées ou qui ne sont même pas à jour. Autrement dit, des informations manquantes, incomplètes, erronées... **Il est donc facile de comprendre qu'il n'est pas possible de prendre des décisions éclairées, de prédire, de gouverner ou de sécuriser avec des données de mauvaise qualité et en faible quantité.**

Comment m'assurer que j'exploite bien les données à valeur ajoutée de mon organisation? **Et comment m'assurer de leur qualité dans un contexte de Cloud, de SaaS, où les systèmes d'informations sont de plus en**

plus ouverts, interdépendants et les données de plus en plus éparpillées? Heureusement, il existe des outils permettant de répondre à cette problématique et de mettre en œuvre une stratégie de Data Quality, basée sur un catalogage des données le plus complet possible. Cette démarche de qualité répond à de nombreux enjeux:

- ✓ **La localisation du patrimoine data**, pour une vue 360° de toutes vos données
- ✓ **Les référentiels**, pour centraliser vos données stratégiques et mieux les partager avec votre écosystème
- ✓ **L'automatisation**, pour le contrôle de cohérence de vos données avec des règles métier, et assurer la qualité par le nettoyage systématique
- ✓ **Le droit à l'information**, pour identifier les données collectées sur un collaborateur ou un client
- ✓ **L'anonymisation des données**, pour assurer la confidentialité
- ✓ **L'Intelligence Artificielle, la Business Intelligence**, pour assurer l'innovation et maintenir un niveau concurrentiel élevé

Nous allons vous expliquer dans ce guide les bases d'une stratégie de Data Quality efficace et comment initier cette démarche grâce à la cartographie de votre patrimoine Data.

175 zettaoctets
dans le monde en 2025*

1 zettaoctets
= **1Md téraoctets***

0,5%
de ces données sont
actuellement analysées*

80%
sont des données
non structurées*

*Source: IDC



Notre définition

« La qualité des données, ou Data Quality, désigne la mesure dans laquelle les données sont précises, complètes, fiables, pertinentes et à jour. Elle évalue également la cohérence des formats de données, leur accessibilité et leur pertinence pour les besoins spécifiques des utilisateurs ou des systèmes qui les exploitent. »

Une haute qualité des données est donc essentielle pour une prise de décision éclairée, l'analyse précise et l'optimisation des processus des entreprises et des organisations.



Les leviers pour répondre à cet enjeu

« La Data Quality est le sujet de tous. La sensibilisation de chaque acteur (producteur et consommateur de données) est primordiale, ainsi que la mise en place de procédures qui pérennisent cette gouvernance des données. »

Mais au regard de l'explosion de la volumétrie de ces données, l'implication de tous n'est pas suffisante, les outils deviennent indispensables. À condition qu'ils communiquent efficacement entre eux et que l'accès à l'ensemble des données soit possible. **Donc pour que la démarche soit efficiente, une première étape est impérative: cartographier son patrimoine data.**



Les facteurs impactant

« Le plus souvent, les actions pour corriger la qualité des données sont effectuées lors des traitements. Car si les métiers savent analyser leur donnée, ils ne sont pas directement impliqués dans leur nettoyage, cette tâche étant encore souvent dévolue à l'équipe IT qui n'a pas la même connaissance du besoin et ne peut intervenir en amont. »

Et avec les changements et évolutions des équipes IT, il est difficile de maintenir les spécificités et de mémoriser les règles définies. **De plus, si les équipes métiers montrent une implication restreinte, l'organisation va connaître une récurrence des problèmes de qualité des données, et c'est un cercle vicieux assuré.**

Data Quality, qu'est-ce que c'est ?

Comment initier une démarche de Data Quality ?



Quels sont les outils pour soutenir cette démarche de Data Quality ?

Pour assurer une qualité optimale des données, il est essentiel de disposer des bons outils. Le tableau suivant montre la corrélation entre les différents usages de la Data Quality et les solutions disponibles pour chaque étape de cette démarche.

FONCTIONNALITÉ	OBJECTIF	EXEMPLES
Gestion des données maîtres	Gérer les données maîtres de manière centralisée et de les synchroniser entre différents systèmes et applications.	Création d'un référentiel de clients unique, synchronisation des données produits entre les différents canaux de vente.
Analyse de la qualité des données	Vérifier la qualité des données en détectant les erreurs, les incohérences et les données manquantes, et de les corriger.	Détection des doublons dans les données clients, correction des adresses postales incorrectes.
Sécurité des données	Protéger les données contre les risques de sécurité, tels que les violations de données, les attaques de pirates informatiques, etc.	Contrôle d'accès aux données, cryptage des données sensibles.
Mise en conformité	Vérifier que les données sont utilisées de manière conforme aux lois et aux réglementations en vigueur.	Suivi des exigences RGPD, mise en place de contrôles pour garantir la confidentialité des données.
Solution de surveillance et d'analyse	Surveiller les données et les processus de gouvernance, d'identifier les lacunes et d'améliorer en conséquence.	Surveillance des performances des données, identification des risques de sécurité.
Data Discovery	Découvrir les données qui se trouvent dans les différents systèmes et applications et de les indexer pour une recherche plus facile.	Identification des données sensibles, recherche des données nécessaires pour un projet particulier.
Data Catalog	Référencer et documenter le patrimoine de données de l'organisation.	Création d'un catalogue des données disponibles, documentation des métadonnées des données.
Data Lineage	Analyser dynamiquement les Systèmes d'information hétérogènes et détecter les usages des données.	Suivi de la provenance des données, identification des impacts d'une modification des données.
Partage de données	Centraliser des données APIfiées pour favoriser la ré-utilisation.	Création d'un portail de données pour les employés, partage des données avec des partenaires externes.
Transformation des données (ETL, ELT, ESB, APIM)	Transformer des données brutes en données exploitables.	Nettoyage des données, transformation des données pour l'analyse, création de pipelines de données.
Visualisation et de reporting	Générer différents types de rapports.	Création de tableaux de bord, génération de rapports PDF, visualisation des données sur des cartes.



Best of breed ou solution unifiée, quelle stratégie adopter ?

Auparavant la réponse à cette question était facile: il n'existait pas de solution unifiée capable de couvrir l'ensemble des enjeux d'une stratégie de Data Quality. Il fallait donc agréger différents outils et réussir à les faire communiquer entre eux. **Avec l'alliance de Phoenix et de MyDataCatalogue, Blueway est capable de fournir dans une approche unifiée tous les outils nécessaires à cette stratégie.**

« Qu'il s'agisse de faire du profilage, du dédoublonnage, du nettoyage, de l'homogénéisation, de la duplication, de l'enrichissement... Toute cette démarche itérative de suivi de qualité de la donnée peut être traitée avec l'ensemble des fonctionnalités des modules proposés aujourd'hui par Blueway. »

3 conseils pratiques pour une mise en œuvre efficace de votre stratégie de Data Quality

1

*« Pour soutenir cette démarche de qualité de la donnée, une campagne de sensibilisation menée en amont et portée par un sponsor issu de la direction est bienvenue. **Il convient de mettre en place un plan de communication adapté avec des actions de formation et des supports clairs et attractifs pour sensibiliser les collaborateurs et les inciter à adopter les bonnes pratiques.** »*

2

*« La qualité des données est un sujet transverse, elle ne doit pas être la préoccupation des seules équipes Data ou IT. **Il faut donc réimpliquer les métiers et les accompagner directement dans leurs applications métiers pour faire du contrôle et de la validation de la cohérence de leurs données.** Ils ont désormais des outils à disposition, comme MyDataCatalogue, pour se saisir de cet enjeu. »*

3

*« Les contrôles de qualité tout au long du cycle de vie des données sont essentiels. Car si une donnée est de mauvaise qualité à l'origine, elle le sera également à la fin, c'est la notion de Garbage In Garbage Out. **Pour garantir l'efficacité d'une stratégie de Data Quality, il peut donc être intéressant de mener des ateliers avec les métiers pour positionner des règles de suivi de qualité directement dans les applications métiers.** »*

Retour d'expérience : le ROI de la qualité



Secteur public

Cas d'usage #1:

Les données concernant un tiers bénéficiaire d'une aide régionale se trouvent à la fois dans un outil de gestion des aides, dans un outil de gestion financière et dans d'autres applications... **sans que ces outils ne puissent communiquer entre eux**. Ainsi, si l'adresse du tiers est erronée, une modification est faite dans l'outil de gestion des aides, mais la remédiation s'arrête là. Les autres systèmes n'ayant pas connaissance de cette modification, les erreurs continuent : retour de courrier de mandatement de l'agent en charge de la gestion financière par exemple, qui doit appeler l'agent gestionnaire des aides pour obtenir la bonne information, ainsi de suite... **Pour une seule erreur de cohérence, l'estimation des remédiations est alors estimée à 1 000 euros par tiers!**



Secteur bancaire

Cas d'usage #2:

Pour définir leur niveau de risque et les règles de conservation à appliquer à des données non structurées dans le domaine bancaire, il faut trouver un outil qui puisse effectuer un scan des serveurs selon des critères de gestion spécifiques, remonter les alertes, jusqu'à effectuer la purge et la mise en conformité de serveurs bureautiques. **La mise en œuvre d'une solution de data cleaning a rendu possible une libération de 95% des espaces disques utilisés sur le périmètre ciblé.**

Synthèse :

les 4 points clés à retenir pour assurer la qualité des données

1 La découverte

→ Connaître et identifier

2 L'analyse

→ Localiser et comprendre

3 La classification

→ Inventorier et trouver

4 La visualisation

→ Maîtriser et sécuriser

Nos conseils pour choisir la solution la plus adaptée

Avant de vous lancer, il est important de tenir compte de vos besoins, mais aussi de la maturité de votre stratégie data pour choisir la bonne solution de catalogage de données. Posez-vous ainsi les questions suivantes :

- 1 Quels sont vos objectifs en matière de Data Quality ?
- 2 Quelles sont vos ressources disponibles ?
- 3 Quelles sont vos exigences en matière de sécurité et de gouvernance des données ?

Mettez ensuite en regard des réponses apportées les points clés suivants pour vous orienter vers la solution la plus adaptée :

- ✓ **Facilité d'utilisation et d'intégration :** La solution doit être facile à utiliser et à intégrer à votre environnement informatique existant. Elle doit également être accessible aux utilisateurs non-techniques.
- ✓ **Gouvernance des données et sécurité :** la sécurité et la confidentialité de vos données doivent être garanties. La solution doit également vous permettre de définir des règles et des politiques d'accès aux données.
- ✓ **Évolutivité et flexibilité :** La solution doit pouvoir évoluer avec vos besoins en matière de données, tout en étant flexible et adaptable à vos exigences spécifiques.
- ✓ **Support et formation :** l'éditeur doit proposer un support technique de qualité. Il doit également proposer des formations pour vous aider à utiliser la solution de manière optimale.



Passez à l'action avec **MyDataCatalogue**, la solution de catalogage de données qui vous permet de prendre de meilleures décisions basées sur **des données fiables et précises** / Augmenter votre efficacité en **réduisant le temps consacré à la recherche et à la correction des données** / Réduire vos coûts en **minimisant les erreurs et les duplications de données** / **Améliorer la satisfaction de vos clients**, autant en interne qu'en externe.

À propos de Dawizz

Dawizz, une société Blueway, est l'éditeur de la plateforme de catalogage et de cartographie des données MyDataCatalogue.

Déployable en mode SaaS ou on-premise, elle permet grâce à ses algorithmes intelligents, de recenser et de classifier automatiquement et exhaustivement les données du patrimoine informationnel.

MyDataCatalogue est l'un des 5 modules de la plateforme Phoenix de Blueway. Phoenix permet aux équipes IT d'intégrer, de gouverner et de partager leurs données, offrant ainsi une solution globale de Data Management pour repositionner l'humain au cœur du système d'information. Le résultat? Une organisation plus agile, des collaborateurs plus autonomes et des liens plus durables au sein des écosystèmes de ses clients.



Stéphane LE LIONNAIS

Véritable touche-à-tout et entrepreneur dans l'âme, Stéphane possède une expérience dans divers secteurs (métallurgie, plasturgie, menuiserie, puériculture, cosmétique et informatique), dont plus de 15 années dans l'univers de la Data (Responsable de l'innovation autour de la démarche Open Data, Directeur de projet dans l'intégration de solutions ERP, BI...). Il est aujourd'hui responsable de la BU Dawizz chez Blueway.

Pour plus d'informations sur Blueway
et ses solutions, visitez www.blueway.fr

Vous souhaitez mieux maîtriser
votre patrimoine de données ?

Contactez-nous! →

