

# CAHIER N°3 : APPROCHES, PROCESSUS ET ORGANISATIONS POUR DEPLOYER UNE DEMARCHE DE GESTION DE LA QUALITE DES DONNEES

DAMA – ISACA - AFAI  
2024

La gestion de la qualité de la donnée est cruciale pour tout type d'entreprise quels que soient sa taille, son secteur, son emplacement géographique, son modèle économique et organisationnel.

La qualité de la donnée est importante pour :

- ❖ Favoriser une prise de décision éclairée
- ❖ Assurer une conformité réglementaire
- ❖ Améliorer l'efficacité opérationnelle
- ❖ Améliorer la satisfaction des clients grâce à des produits qui se basent sur des données de qualité
- ❖ Monétiser la donnée en répondant aux exigences de qualité de ses consommateurs
- ❖ Acquérir des avantages concurrentiels en maîtrisant les données sectorielles

Bien que les objectifs globaux de la gestion de la qualité des données se rapprochent, les approches, processus et organisations déployées varient d'une entreprise, entité et équipe à l'autre.

Dans ce troisième cahier, nous recensons les différents types d'approches, processus, modèles opératoires que nous observons dans les entreprises afin de dégager des facteurs clés de succès de ce type d'initiatives.

Ce cahier a été rédigé dans le cadre d'un GT animé par DAMA et l'AFAI.

## Question 1 : Quels sont les types d'approches possibles de gestion de la qualité de la donnée ?

Dans cette section, nous présentons diverses approches de gestion de la qualité des données, chacune adaptée à des besoins spécifiques et ayant ses propres objectifs que nous résumons dans le tableau ci-dessous. Ces approches ne sont pas mutuellement exclusives.

Approche	Description
<b>Top Down/ Bottom up</b>	<ul style="list-style-type: none"><li>• Une approche top-down qui consiste à concevoir les processus et les data <i>from scratch</i> à partir d'une définition théorique de ce que devraient être les-dits process et en constituant une version idéale. Cette version idéale pourra être le ciment de la définition d'un langage commun entre les processus métier. Les Master data constituent à ce titre une part importante des données concernées par cette approche. Dans une organisation disposant de plusieurs établissements dotés d'une certaine autonomie, cette approche permet de diffuser un langage commun et de définir des indicateurs de performance transverses entre les sites.</li><li>• Une approche terrain Bottom-up en partant des process et des difficultés engendrées par la non-qualité des données (dysfonctionnements, retards, performance dégradée, rebuts, arrêts de ligne de production, etc.)</li></ul>
<b>Centralisée / Décentralisée / Fédérée</b>	<p>A l'instar de l'approche théorique / Terrain précédente, une alternative peut consister à prendre une perspective de gouvernance et se caler aux modèles définis par DAMA sur la gouvernance de données</p> <ul style="list-style-type: none"><li>• Une approche centralisée où la démarche de l'ensemble de la structure, potentiellement composée de plusieurs établissements, est définie centralement et appliquée localement.</li><li>• Une approche décentralisée où la question de la qualité de données est laissée à l'initiative des sites locaux sans intervention des équipes centrales.</li><li>• Une approche fédérée où la question de la qualité de données est définie au niveau central qui coordonne les différents niveaux locaux responsables eux-mêmes de la qualité de leurs données. Dans cette approche de qualité de données, les responsabilités sont partagées entre le niveau central et local par domaine de données</li></ul>

<b>Backlog priorisé par la valeur métier</b>	<p>Même si cette thématique est récurrente et ne doit pas faire l'objet d'une attention uniquement en début de démarche, elle demeure essentielle et assure la pérennité de l'initiative.</p> <p>Il apparaît essentiel d'initier un backlog des erreurs de qualité de données et de les prioriser en fonction de critères métier.</p> <p>On pourra par exemple considérer comme critère de priorisation, les erreurs portant atteinte à la sécurité des personnes ou des biens, celles préjudiciables à l'exécution des missions de l'entreprise, des fonctions métier, celles qui induisent des retards d'exécution, celles qui dégradent la qualité de la production ou encore celles qui portent atteinte à l'image de l'entreprise.</p> <p>Il est souhaitable que la priorisation du backlog soit réalisée selon un rythme régulier et s'appuie sur les instances de décision mises en place par la gouvernance des données (ou a contrario que les réunions périodiques de priorisation soient également le prétexte pour y traiter également d'autres sujets liés à la gouvernance).</p>
<b>Proactive / Réactive</b>	<ul style="list-style-type: none"> <li>• L'approche réactive est celle qui est souvent considérée en premier lieu et consiste à venir mettre en place une démarche de qualité de données (a minima du cleansing) à la suite de la découverte d'un ou plusieurs incidents de qualité de données ou d'une prise de conscience de l'impact de la non-qualité sur les processus opérationnels</li> <li>• L'approche proactive consiste à adopter une démarche "By Design", c'est-à-dire anticiper les problèmes de qualité de données dès la mise en place d'un nouveau processus métier lors de l'implémentation des outils numériques supportant ce processus métier.</li> </ul> <p>Par extension, on pourra assimiler la résolution d'un problème (au sens ITIL) de qualité de données à une approche proactive (mise en place de contrôle de saisie, etc.)</p>
<b>Offensive / Défensive</b>	<ul style="list-style-type: none"> <li>• L'approche défensive consiste à optimiser les processus métier existants en améliorant leur performance et leur apport de valeur à l'Entreprise et à son écosystème. C'est l'approche la plus répandue dans les entreprises qui disposent d'un patrimoine informationnel important qui a longtemps été considéré comme un flux opérationnel, comme le traitement d'une commande, l'émission d'une facture, la résolution d'un ticket.</li> <li>• L'approche offensive consiste à donner les moyens à la structure (entreprise, entité...), au travers de la qualité des données, de développer de nouveaux assets, offres, produits au bénéfice de cette structure ou de son écosystème. L'IA générative est notamment rendue possible car faisant appel à des jeux de données d'apprentissage de grande qualité. A titre d'exemple, on pourra à ce titre citer des entreprises telles que Palantir, DataRobot ou Zillow qui ont su capitaliser sur des données de très haute qualité pour développer des modèles d'apprentissage ou proposer de nouveaux services.</li> </ul>

**Par typologie de données : Data / Master data / Reference data**

L'approche considérée pour traiter des questions de qualité de données peut également prendre en compte la typologie des données car chaque type de données requiert une approche différente :

- **Données transactionnelles** : Ces données sont généralement descriptives d'un processus métier bien défini (commande, facturation, ordre de fabrication, etc.). Ces données sont généralement produites par une application et sont celles qui circulent globalement le moins dans le système d'information et sont les moins réutilisées.  
Concernant ces données, les critères de fraîcheur, d'exactitude et de validité sont ceux que l'on sélectionne en premier lieu
- **Master data** : Contrairement aux données transactionnelles, les master data sont des données qui sont partagées par plusieurs processus d'un département voire de l'entreprise dans son ensemble. Le client, le produit sont des exemples courants de master data.  
Ces données partagées pouvant à la fois être produites par plusieurs processus/applications et susceptibles d'être modifiées au sein d'une application, les critères respectivement d'unicité et d'intégrité pourront être favorisés.
- **Les données de référence** sont des données statiques ou qui changent très peu au cours du temps. Elles peuvent être également définies par des autorités externes à l'Entreprise et dont l'objectif est de définir des standards.  
On pourra par exemple citer la liste des pays concernés par pas moins de 5 normes ISO ou localement le référentiel des codes postaux. Sur ces données, dont la qualité est principalement portée par l'autorité de laquelle elles émanent, l'attention pourra être priorisée sur la fraîcheur.

## Par Outillage

Une approche lorsque l'on se lance dans une démarche de qualité de données concerne également l'outillage.

Sur ce sujet, la question du "Make or Buy" peut se poser, mais la réponse dépendra des orientations, politiques, stratégies de chacun.

L'acquisition d'une licence d'une application de gestion de la qualité de données ne doit pas pour autant masquer l'effort à fournir par l'entreprise pour exploiter cet outil, en assurer le déploiement et l'adoption et en faire un outil au service de la gouvernance.

Le choix du scénario Make, approche plus complexe, voire risquée, pourra néanmoins s'appuyer sur des briques existantes du système d'information de l'entreprise, telles que l'entrepôt de données, pour construire les indicateurs de mesure de la qualité de données en bénéficiant d'un socle de données transverses.

Qu'il soit sur étagère ou tout droit sorti du fournil des data engineers, on pourra trouver dans un outil de qualité de données tout ou partie des fonctionnalités suivantes :

- Connexion aux applications de l'entreprise
- Définition des critères de qualité de données
- Définition d'un à plusieurs seuils de qualité par attribut ou groupe d'attributs
- Mesure de la qualité de données systématique ou à la demande
- Mise à disposition des jeux de données passantes / non passantes
- Notifications des personnes ressources de l'Entreprises (cf. data gouvernance) habilitées à recevoir et traiter ces notifications
- Capacité à mettre en œuvre des indicateurs et des tableaux de bord sur la qualité de données
- Gestion de la confidentialité et la sécurité des données
- Capacité à communiquer sur l'actualité, la feuille de route, mais aussi sur les succès (et les difficultés) du programme

## Question 2 : Quels sont les processus, organisations et modèles opératoires possibles ?

La qualité de la donnée, telle que définie dans le cahier pratique numéro 1, est étroitement corrélée au contexte et à l'usage des consommateurs de cette donnée. Le processus de contrôle de la qualité de la donnée a pour objectif de servir de façon optimale l'exécution d'un processus au sein d'une organisation.

Dans ce chapitre, nous partageons des éléments pour la définition d'un tel processus dans une organisation, des rôles et responsabilités qui y sont liés ainsi que du mode opératoire. Pour ce faire, nous commencerons par faire quelques rappels.

La qualité de données est **une démarche d'entreprise**. Elle concerne l'ensemble des acteurs de la donnée au sein de l'entreprise mais aussi les acteurs externes. Elle ne peut être définie comme un projet. En ce sens, il est préférable qu'elle soit vue comme une fonction transverse et qu'elle bénéficie d'un financement indépendant des fonctions métiers ou alors intégrée de façon systématique dans les financements métier de la même façon que les fonctions de conformité réglementaire ou de sécurité le sont.

Rappelons qu'**un niveau à 100% de qualité de données n'est pas envisageable**. En revanche, la définition d'un niveau acceptable de qualité par périmètre et la surveillance du maintien de ce niveau attendu permet d'améliorer l'excellence opérationnelle des processus d'entreprise concernés.

De nos jours, il est établi que la donnée est un actif pour une organisation. Son usage et le bénéfice qui en est tiré impliquent la responsabilisation des métiers sur la donnée, sa définition et son cycle de vie, depuis sa production et jusqu'à sa destruction. **Les métiers sont ainsi amenés à intégrer dans leurs feuilles de routes la mise en place et le maintien d'un processus adéquat de contrôle de la qualité**. Ce processus peut induire des changements à la fois applicatifs et/ou purement organisationnels.

### Processus d'amélioration et de maintien de la qualité de la donnée

Ce paragraphe vient en complément du travail de définition du périmètre de données à contrôler et les critères de qualité présentés dans le cahier pratique numéro 2. Dans ce dernier, deux principales approches, par la valeur et par les produits finis, ont été abordées. Dans ce cahier, nous nous intéresserons davantage à l'approche hybride combinant données et processus organisationnels.

Rappelons qu'un processus de qualité de la donnée est empirique et nécessairement itératif. Chaque itération apporte un niveau d'affinage supérieur et des critères de sélection et de priorisation adaptés. Dans une approche hybride du processus de qualité, une connaissance à la fois des données de l'entreprise ainsi que des processus organisationnels est essentielle.

On s'intéressera dans un premier temps à définir les principaux concepts, notions ou objets métier manipulés dans le cadre du modèle d'affaires d'une organisation. À titre d'exemple, ces notions peuvent inclure les notions de client, site, produit, patient, employé, actionnaire etc. dépendamment du domaine d'affaire de l'organisation.

Dans un second temps, toujours dans un mode itératif, on s'intéressera aux processus métier (business process) manipulant ces concepts et permettant à l'organisation d'accomplir sa mission. Ceux-ci peuvent impliquer, par exemple, des processus opérationnels contribuant à la chaîne de valeur de l'entreprise (traitement des commandes client, développement de produits etc.), des processus de support (contrôle de gestion, RH etc.) et des processus de management (gouvernance, planification stratégique et budgétisation etc.) selon la mission de l'entreprise.

Enfin, on s'intéressera aux systèmes et aux acteurs permettant d'opérer les notions et les processus pour accomplir les missions de l'organisation.

La combinaison de prismes : concepts, processus, systèmes et acteurs permet d'identifier les flux de données dans l'entreprise et de déterminer la responsabilité des métiers sur la donnée, sa définition, son usage.

De cette combinaison découleront les contrôles de qualité adéquats à implémenter. Elle permettra aussi de prioriser les données à contrôler selon la criticité, via une approche par la valeur ou par le produit fini et de cibler les actions correctives (acculturation, changement de processus, évolutions applicatives etc.) tel que décrit dans le mode opératoire plus loin dans ce chapitre.

## Organisation

Dans ce paragraphe, nous identifions des éléments type d'organisation ainsi que les rôles et responsabilités pour opérer le processus de qualité de la donnée.

Notons d'abord que ce processus sera facilité par l'existence d'une instance décisionnelle centralisée ou fédérée de management de la donnée. Elle peut prendre différentes formes (Data Office, Data Council etc.) et a pour rôle d'influencer le traitement de la donnée dans l'organisation et la mise en place des fondations pour une gouvernance appropriée incluant les questions de l'acculturation et vulgarisation, l'outillage et catalogage, la mise en place d'entrepôt de donnée, la gestion des Meta data et l'architecture.

Cette instance (Data Office, Data Council etc.) pilote notamment un comité décisionnel spécifique à la qualité de la donnée chargé de la priorisation des périmètres de la donnée à traiter, de la définition des dimensions de la qualité de la donnée, de l'affinage des métriques et des règles d'agrégation. Il repose sur des rôles au sein de l'organisation à la fois du côté métier et du côté IT.

Du côté métier, on identifiera principalement les propriétaires de la donnée (Data Owner) et les propriétaires de processus métier (Process Owner). Le but est d'apporter la connaissance fonctionnelle sur les processus producteurs et consommateurs de la donnée, sur la définition de la donnée, les règles fonctionnelles qui y sont reliées afin d'être en mesure de définir les dimensions adéquates de qualité, les niveaux attendus et d'impulser les changements organisationnels nécessaires pour implémenter des correctifs métier pour redresser la qualité.

Du côté IT, on identifiera principalement les intendants de la donnée (Data Custodian) et les responsables applicatifs. Le but ici est d'apporter la connaissance technique sur les systèmes producteurs et consommateurs de la donnée, la connaissance des flux de transformation et du data lineage et d'implémenter les règles et mesures de qualité conjointement identifiées par les acteurs métier.

On identifiera également les rôles d'urbaniste et d'architecte de la donnée qui font la jonction entre le métier et l'IT et qui portent la vision de l'entreprise sur les applications maître de la donnée et les processus.

## Mode Opérateur

Dans ce paragraphe, nous proposons un mode opératoire du processus de contrôle de la qualité de la donnée. Ce mode est décrit sous la forme d'un processus cyclique passant par les étapes suivantes :



## Question 3 : Comment déployer ces processus, organisations et modèles opératoires ?

Une initiative de gestion de la qualité des données est bien plus qu'un simple processus technique ; c'est une transformation organisationnelle qui exige une vision stratégique, une collaboration multidisciplinaire et un engagement continu. Au cœur de ce défi se trouve la nécessité de déployer efficacement les processus, les organisations et les modèles opérationnels nécessaires pour garantir la qualité des données à chaque étape de leur cycle de vie.

Nous avons **identifié onze (11) points clés** pour déployer avec succès une initiative de gestion de la qualité des données au sein d'une entreprise. Ces points clés ne sont pas cités par ordre chronologique. Chacun de ces points peut être repris lors du déploiement d'un programme de Data Quality selon l'opportunité et le besoin qui se présentent.

### 1. Comprendre les besoins métier, leurs difficultés et leurs objectifs

Le point de départ de toute initiative, projet ou démarche de gestion de la qualité des données est la compréhension des besoins des métiers. Cette activité consiste à connaître et à comprendre les éléments dont les métiers ont besoin (intégrité, exhaustivité et exactitude de la donnée) pour améliorer leur prise de décision, assurer la continuité de leurs opérations et répondre aux exigences réglementaires.

- Les métiers sont-ils conscients des enjeux de qualité de données sur leur périmètre ?
- Quelles sont les difficultés qu'ils rencontrent au quotidien par rapport à la qualité des données ?
- Quels sont les fichiers qu'ils doivent corriger au quotidien ?
- Combien de perte de temps cela génère-t-il ?
- Quel est l'impact d'une mauvaise qualité des données sur leur activité ?
- L'impact est-il faible, moyen ou élevé ?
- De quel ordre est l'impact : réputationnel, social, financier, satisfaction du client, opérationnel ?
- Quelles sont les causes racines de ces problèmes de qualité ?

Il pourrait être utile de s'inspirer du diagramme de Fishbone pour structurer les causes racines et les actions de remédiation prioritaires pour les métiers.

Il est également important d'identifier les besoins en termes de standards de qualité : quels sont les besoins de normes de qualité pour chaque processus métier ? quelles sont les normes communes entre les différents consommateurs de la donnée ?

Ce travail d'alignement garantit une compréhension commune des attentes en matière de qualité entre producteurs et consommateurs de la donnée.

#### **Exemple de point de départ :**

- Reporting à la direction non fiable
- Expérience client insatisfaisante
- Réduction de coût, amélioration performance commerciale

- Exigences réglementaires (BCBS239)
- Initiative spontanée locale

### **Exemple d'un déroulé d'atelier (reporting non fiable)**

- *Quels sont les indicateurs qui vous semblent erronés ? Pourquoi ?*
- *Quelles sont les explications potentielles ?*
- *Quelles sont les données sources ? Sont-elles fiables ? Sont-elles compréhensibles et disponibles en temps opportun ?*
- *Quelles sont les personnes intervenant dans l'élaboration du reporting ? Quel est leur rôle ?*
- *Quels sont les outils utilisés pour mettre en place le reporting ?*
- *Qui peut appuyer votre démarche au sein de l'entreprise ?*

L'entretien permettra de faire ressortir les points de douleur. En effet, un point de douleur peut se situer au niveau de l'intégrité des données recueillies auprès des processus métiers ou se matérialiser par une absence d'outils ou d'automatisation des processus.

## **2. Etablir des critères de priorisation des initiatives et problèmes de qualité de la donnée**

Déployer c'est aussi prioriser. Il est important d'adopter une approche pragmatique lorsque nous déployons une démarche de qualité de la donnée. Dans un contexte de ressources et de temps limités, les sujets ne peuvent pas être traités tous en même temps, surtout lorsque les sujets sont traités en central et que le Data Office en question n'a pas assez de ressources pour résoudre les problématiques transversales identifiées.

### **Comment prioriser ?**

- Si nous sommes dans une approche défensive, les sujets réglementaires seront largement priorisés par rapport aux autres problématiques
- Sinon dans une approche offensive, il est recommandé, grâce à une matrice de priorisation, d'évaluer :
  1. L'impact et la criticité du problème sur le(s) processus métier(s)
  2. Le risque associé à traiter ou non le problème
  3. Les ressources et coûts nécessaires pour corriger ou prévenir ces problèmes
  4. La complexité de résolution
  5. Les gains à corriger ou prévenir le problème
  6. L'alignement stratégique avec les objectifs de l'entreprise : si le problème impacte un chantier prioritaire, il est recommandé de le prioriser

Cette matrice de priorisation permettrait de mitiger les risques et résoudre avec pragmatisme les problèmes identifiés.

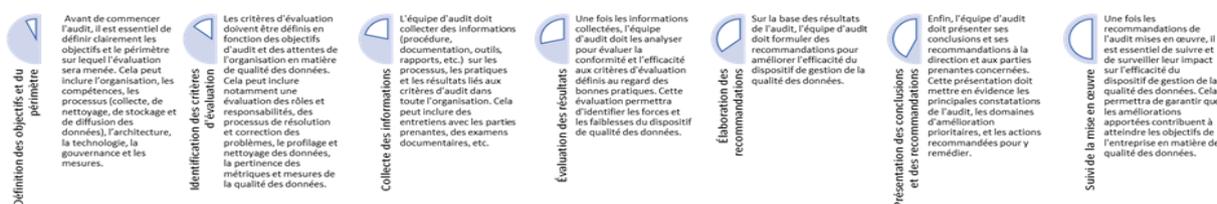
## **3. Faire un audit et un profiling sur un périmètre clé**

La mise en œuvre de tout dispositif de gestion de la qualité des données doit être inscrite dans un processus d'amélioration continue afin de répondre de la meilleure des manières aux exigences du dispositif. Pour garantir l'efficacité et l'évolution continue d'un dispositif de gestion de la qualité des

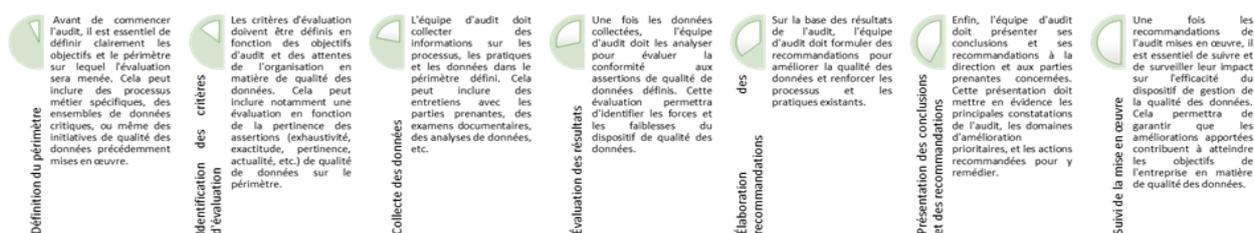
données, un audit indépendant représente une démarche essentielle. Il permet d'identifier les axes d'amélioration, de justifier les investissements nécessaires et d'assurer l'alignement sur les objectifs stratégiques de l'organisation.

Deux types d'audits indépendants se distinguent particulièrement dans ce processus : l'audit de l'efficacité du dispositif de gestion de la qualité des données et l'audit de la qualité des données sur un périmètre clé.

L'audit de l'efficacité du dispositif de gestion de la qualité des données vise à évaluer de manière impartiale les processus, l'organisation et le modèle opératoire mis en place dans le cadre de la gestion de la qualité des données au sein de l'entreprise. Cette évaluation rigoureuse permet de mesurer la performance et l'impact des initiatives mises en œuvre, conformément aux normes et standards de qualité de données. La réalisation de cet audit passe par les étapes suivantes :



D'autre part, l'audit de la qualité des données sur un périmètre clé se concentre sur l'évaluation de la qualité des données gérées dans des périmètres spécifiques de l'entreprise. En s'appuyant sur des assertions de qualité de données pertinentes, cet audit identifie les forces et les faiblesses de la qualité des données, contribuant ainsi à une amélioration continue. La réalisation de cet audit passe par les étapes suivantes :



La composition de l'équipe chargée de mener ces audits revêt une importance capitale. Composée d'experts en qualité des données et d'auditeurs internes ou externes, cette équipe doit posséder une connaissance approfondie des normes et des bonnes pratiques en matière de gouvernance et de qualité des données.

## 4. Créer une équipe multidisciplinaire

La qualité des données est un sujet pluridisciplinaire. Il concerne tous les métiers de l'entreprise.

**Les métiers en front** : producteurs et consommateurs de la donnée, ils identifient des problèmes, ont des exigences de qualité et savent évaluer l'impact d'une non-qualité sur leur activité et donc sur l'entreprise (impact réglementaire, impact client, impact produit)

**Les métiers supports et RH** lorsque nécessaire : le maintien d'une donnée RH à jour dans le cadre de la gestion du cycle de vie d'un collaborateur est importante (impact juridique, financier, salarial, organisationnel)

**Les équipes IT et la DSI** : producteurs et consommateurs de données, ils doivent trouver les solutions de collecte, stockage et transfert de données les plus adaptées pour répondre aux exigences de qualité. Ils sont impactés par la mauvaise qualité lorsqu'ils doivent croiser des données ou assurer l'interopérabilité entre les systèmes.

**Les équipes juridiques** : lorsqu'il s'agit d'exigences réglementaires, RGPD ou informations commercialement sensibles impactant la qualité des données, il est important d'impliquer la direction juridique pour s'assurer que la qualité est conforme aux obligations réglementaires

**Les Architectes** : selon le sujet, il est important d'inclure les architectes d'entreprise. Par exemple, si nous détectons un problème de qualité des données, savoir quels systèmes applicatifs et processus métiers sont impactés est primordial.

**Chief Data Officer** Si le problème est critique ou l'initiative est clé pour atteindre les objectifs stratégiques, il est important que le Chief Data Officer soit engagé, impliqué et avec une posture engageante auprès des parties prenantes. Le CDO doit avoir une vue sur le tableau de bord de la qualité de la donnée afin de prioriser les actions et les ressources.

**Les Data Managers et analystes de la qualité de la donnée** : Les Data Managers vont assurer une vue d'ensemble sur ces différentes problématiques et proposer des plans de remédiation tout en prenant en compte les exigences pluridisciplinaires. Ils vont également s'appuyer sur les analystes pour prendre des décisions basées sur une qualification chiffrée des problèmes identifiés. Les Data Managers vont également suivre de bout-en-bout la démarche : depuis le cadrage jusqu'à l'opérationnalisation des actions de prévention à la source.

**AI Officer** si le problème de qualité impacte une intelligence artificielle développée ou acquise.

## 5. Identifier les rôles et responsabilités manquants nécessaires au déploiement

Lors de la constitution de l'équipe multidisciplinaire et du déploiement de la démarche par BU, par processus métier ou par domaine de données, un écart peut être observé entre les rôles et responsabilités actuels et ceux nécessaires pour entreprendre les actions de remédiation de la qualité des données

Dans ce cas, il est nécessaire de formaliser cet écart et de préciser les évolutions de rôles et de responsabilités souhaitées. L'implication du management et des équipes RH est primordiale pour appuyer la démarche et sensibiliser les sponsors.

Dans un modèle de gouvernance décentralisé où chaque BU gère en autonomie la qualité des données, la clarification des rôles est d'autant plus importante pour pouvoir suivre l'avancement des initiatives et alerter lorsque les initiatives ne sont pas sponsorisées.

Dans un modèle centralisé, il est essentiel de clarifier le rôle du Data Office : voudrait-on un Data Office qui fait ou qui s'assure que c'est fait ? Ceci a des impacts sur le dimensionnement du Data Office.

Dans tous les types d'organisation, **les personnes identifiées devraient être capables de dialoguer avec des fonctions transverses** pour :

1. Harmoniser les standards et les normes de qualité
2. Collecter les besoins des autres équipes
3. Communiquer ses exigences aux autres équipes, etc.
4. Communiquer sur l'amélioration continue
5. Assurer en engagement des sponsors

## 6. Identifier les sujets qui doivent être traités en central et ce qui serait mieux traité localement

Certaines problématiques locales peuvent être résolues avec une solution locale si la cause racine du problème est locale : nouvelle organisation locale de la collecte des données, vérification de la conformité avec les standards par un agent local, etc. Exemple : données opérationnelles spécifiques à un site de production.

D'autres problématiques locales ne peuvent pas être résolues localement de manière silotée car la cause racine du problème de qualité est globale. Il est donc important d'instaurer un système de gouvernance central ou fédéral sur ces sujets. Exemples : manque de standards sur des données Master & Reference data, données Client, données financières consolidées.

Lorsque le modèle de gouvernance des données est décentralisé, le traitement des sujets transverses est d'autant plus complexe et nécessiterait des pratiques managériales spécifiques : mise en place des groupes de travail, des communautés, etc. pour sensibiliser à une gestion commune de la qualité des données partagées.

La décision de centraliser ou de localiser la gestion de la qualité des données pour un type de données précis doit s'appuyer sur une connaissance approfondie des flux de données et de leur utilisation par les différentes unités de l'organisation. Les critères d'identification principaux sont l'importance stratégique et la transversalité de l'utilisation des données. Des considérations géopolitiques peuvent également entrer en ligne de compte et amener à une solution hybride de centralisation par zones géographiques.

## 7. Établir des normes standards de qualité

Il est essentiel d'établir des **normes de qualité de données communes et cohérentes** qui prendront en compte les besoins métier et les exigences réglementaires, **surtout sur les données clés de l'entreprise**.

La communication et l'application des normes et standards devrait se faire dès le lancement des projets : Data Quality by design. Les standards doivent néanmoins être régulièrement adaptés et mis à jour pour s'assurer qu'ils restent pertinents face aux évolutions du marché et des technologies. La mise en place d'audits internes ou externes réguliers peut favoriser la conformité aux standards.

## 8. Définir des mécanismes d'escalade des problèmes de qualité des données

L'escalade des problèmes de qualité de la donnée est un sujet organisationnel clé et doit être défini dans la phase de cadrage de chaque projet, initiative ou programme de qualité de la donnée. **Pourquoi mettre en place des mécanismes d'escalade ?**

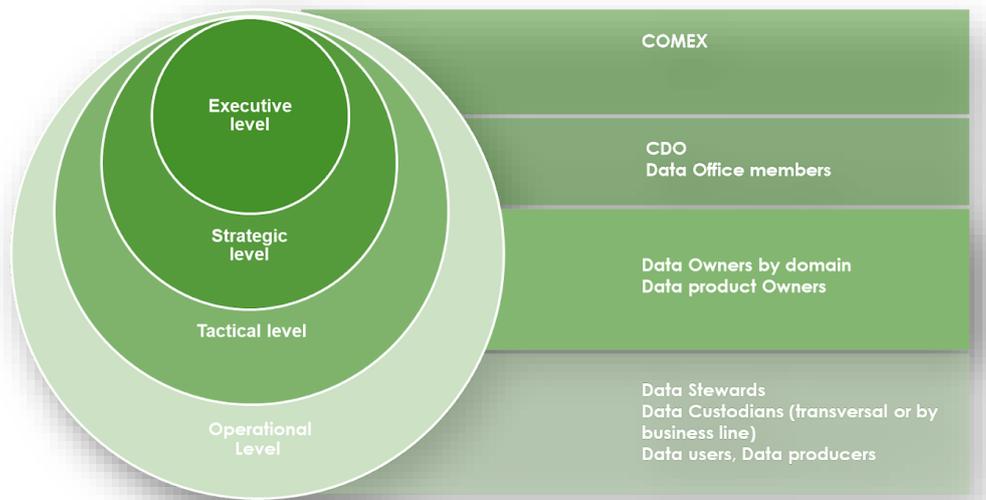
Des mécanismes d'escalade permettent de garantir que les problèmes de qualité des données sont identifiés et résolus de manière rapide et efficace. Ils définissent un processus clair pour signaler, évaluer et corriger les problèmes, en s'assurant que les problèmes critiques sont traités en priorité.

Nous pouvons distinguer six (06) éléments clés d'un mécanisme d'escalade des problèmes de qualité des données :

- **Définition des niveaux de gravité** : Classer les problèmes de qualité des données selon leur impact potentiel sur l'entreprise. Cela permet de prioriser les efforts de résolution.
- **Désignation des responsables** : Définir clairement qui est responsable de la résolution des problèmes de chaque niveau de gravité garantit que les problèmes ne tombent pas entre les mailles du filet.
- **Processus de signalement** : Mettre en place un moyen simple et accessible pour signaler les problèmes de qualité des données peut inclure des formulaires en ligne, des boîtes de réception dédiées ou des points de contact spécifiques.
- **Enquête et analyse** : Établir un processus d'enquête approfondie pour identifier la cause profonde des problèmes de qualité des données. Cela permet de mettre en place des solutions durables.
- **Actions correctives et préventives** : Mettre en œuvre des actions correctives pour résoudre les problèmes identifiés et des actions préventives pour empêcher leur récurrence.
- **Suivi et communication** : Suivre l'évolution des problèmes de qualité des données et communiquer les résolutions aux parties prenantes concernées.

La mise en œuvre effective de ces éléments d'escalade de problèmes permet l'amélioration de la qualité des données, la réduction des erreurs et des anomalies dans les données, ce qui conduit à une meilleure prise de décision. Elle permet également l'augmentation de l'efficacité ainsi que la réduction des coûts.

Exemple : Escalade des problèmes de qualité dans un modèle de gouvernance centralisé ou fédéral



Dans l'idéal, la majorité des problématiques devraient être traitées au niveau opérationnel. L'escalade au niveau tactique ou stratégique se fait dans un deuxième temps lorsqu'un arbitrage est nécessaire.

A noter que dans un modèle de Gouvernance décentralisé, il est important de mettre en place une **gouvernance collaborative** et un cadre global clair et partagé afin de partager les bonnes pratiques (fertilisation croisée) et d'assurer une adhésion des équipes même si elles gèrent en autonomie le sujet pour s'assurer qu'elles participent de manière active à identifier les sujets transverses et communs : comités cross-BU et groupes de travail.

## 9. Mesurer et communiquer les progrès

Pour instiller une culture "qualité des données" dans l'entreprise, il est primordial de mesurer et de communiquer les progrès pour renforcer la démarche, et maintenir sa visibilité sur le long terme.

La communication des progrès est essentielle à tous les niveaux, de l'opérationnel au top management, et implique des équipes Data, métiers ou IT. Elle peut s'effectuer via divers canaux tels que le catalogue de données d'entreprise, les canaux de communication internes habituels, les comités projet ou de pilotage, les cercles de données, ainsi que lors de réunions individuelles ou de comités stratégiques et opérationnels.

Une communication efficace pourra se faire notamment par un tableau de bord visualisant l'amélioration de la qualité des données depuis le lancement du programme, avec des annotations expliquant les initiatives spécifiques ayant conduit à cette amélioration, comme la formation des utilisateurs aux meilleures pratiques de saisie de données. L'inclusion d'indicateurs de qualité des données dans des rapports existants, tels que des rapports financiers ou de performance opérationnelle, pourra également renforcer cette communication.

Il est important d'établir un plan de communication clair : message clé, audience, canal, fréquence et de le mettre en place avec l'appui des parties prenantes qui sont motrices de la démarche.

Pour plus de détails sur ces tableaux de bord, veuillez consulter les chapitres 5 et 6 du cahier pratique n°2 "points de repère clés pour mesurer la qualité des données".

Au lancement de la démarche, la mesure des progrès risque d'être réactive et pourra devenir proactive au fur et à mesure de la montée en compétence et en maturité des parties prenantes (à l'image de tous les produits et services d'une entreprise).

## 10. Former, sensibiliser, faire monter en maturité

Pour déployer efficacement un programme de qualité au sein de l'organisation, il est essentiel de sensibiliser, de former et de faire évoluer la maturité sur les sujets liés à la qualité des données. Ceci permet en particulier de passer progressivement d'un mode réactif à un mode proactif.

La sensibilisation à la qualité des données est nécessaire pour lever certains obstacles au déploiement. L'amélioration de la qualité des données peut, par exemple, être perçue comme une tâche secondaire, non prioritaire. Dans d'autres cas, il peut y avoir une tendance au déni ou une méconnaissance de sa valeur ajoutée ("L'entreprise réalise déjà des bénéfices confortables, alors pourquoi devrions-nous nous donner tant de peine?"). Identifier et comprendre ces obstacles est important pour faciliter des interventions efficaces afin d'initier ou de renforcer la sensibilisation à la qualité des données.

Pour commencer, il est crucial de sensibiliser la direction et les parties prenantes de l'organisation à l'importance de la qualité des données. Cela peut être réalisé en organisant des sessions présentant des études de cas et des exemples concrets illustrant les avantages pour l'organisation. On communiquera également sur les effets néfastes liés à l'absence de qualité de données : perte de temps, d'argent, de réputation et de motivation des employés, impact sur l'environnement, etc. On pourra également évoquer la règle des 1-10-100 qui stipule que le coût de correction des erreurs de données augmente exponentiellement selon le moment de leur détection : 1 unité lors de la saisie, 10 unités lors de la vérification et 100 unités lorsqu'elles impactent les processus métier



Ensuite, il est essentiel d'investir dans la formation et le développement des compétences pour tous les membres de l'organisation. Cela peut inclure des formations sur les concepts de base, les meilleures pratiques de collecte, de gestion et d'analyse des données, ainsi que l'utilisation d'outils de gestion de la qualité des données.

 <p><b>Guide to Proactive Data Quality</b></p> <p><b>Target Audience</b>   Everyone working with data</p> <p><b>Duration</b>   45 minutes</p>	<p><b>Training Highlights</b></p> <ul style="list-style-type: none"> <li>• Raise awareness of data quality in our daily activities</li> <li>• Foster a data quality culture across</li> <li>• Mindset change: inspire colleagues to become data quality champions</li> <li>• To introduce how data quality is measured in ZEISS</li> </ul>
 <p><b>Data Quality Dashboard for Data Owners &amp; Stewards</b></p> <p><b>Target Audience</b>   Data Owners &amp; Data Stewards</p> <p><b>Duration</b>   45 minutes</p>	<p><b>Training Highlights</b></p> <ul style="list-style-type: none"> <li>• How to identify critical attributes and define data quality rules</li> <li>• Features of the data quality dashboard</li> <li>• Manual and automated data quality cleansing approaches from the dashboard</li> <li>• A practical demonstration of using the dashboard to support the data cleansing process.</li> </ul>

Enfin, il est crucial de promouvoir une culture axée sur la qualité des données au sein de l'organisation, en reconnaissant et récompensant les efforts visant à améliorer la qualité des données - via les réseaux sociaux d'entreprise, par un système de prime - et en encourageant la collaboration et la communication entre les différents départements impliqués tout au long du cycle de vie des données, depuis leur production jusqu'à leur utilisation par les différents acteurs.

Une communauté de pratique peut être établie, utilisant par exemple un canal de conversation dédié à la qualité des données. Cela permettrait aux experts et aux novices d'échanger des connaissances et de progresser ensemble. Des événements ponctuels tels qu'une "journée de la qualité des données" peuvent être organisés, offrant une plateforme pour les intervenants internes et externes afin de partager leurs expériences et enrichir la vision de l'entreprise.

**Welcome to Data Quality Day II**

**Today's Agenda**



**03. 11. 2023**

13:30-13:50 pm	● <b>Opening Keynotes – Introduction to Data Quality at the Company</b> - Mr A., Head of <i>BusinessSupportingFunction_xy</i>
13:50-14:30 pm	● <b>Data Quality Drives Business Opportunities</b> - Mr B., Area Sales Representative <i>StrategigBusinessUnit_a</i>
14:30-15:05 pm	● <b>Data Quality: How we do it at <i>ExternalCompany_1</i></b> - Mrs C, Lead Data Scientist at <i>ExternalCompany_1</i>
15:05-15:25 pm	● <b>COFFEE BREAK</b>
15:25-16:00 pm	● <b>Data Quality @<i>ExternalCompany_2</i></b> - Mr D., Vice President Data & Analytics Governance at <i>@ExternalCompany_2</i>
16:00-16:30 pm	● <b>Understand and Prepare your Data - A Quick Guide to Data Readiness</b> - Mr E., Senior Data Quality Analyst at <i>BusinessSupportingFunction_xy</i>
16:30-16:45 pm	● <b>Updates from the Enterprise Data Quality Team</b> - Mrs F., Data Quality Analyst at <i>BusinessSupportingFunction_xy</i>

**La Fresque de la Qualité** développée par DAMA France peut apporter une approche ludique à cet effort d'acculturation à la manière de la fresque climatique. L'objectif est de sensibiliser à la qualité des données et à ses enjeux. Il est proposé un serious game collaboratif autour d'une trentaine de cartes où les participants dessinent une fresque reliant des concepts représentés par des cartes.

### **Une gestion de la qualité réactive versus proactive**

- **Mode réactif**

**Le focus est mis sur l'initiation et la** sensibilisation à la qualité des données. On s'attachera à combattre les idées fausses sur la nature même de la data - *insister sur sa réutilisabilité, sa non-exclusivité à un projet ou à une utilisation spécifique (kidnapping des données)* - et un éventuel fatalisme sur l'amélioration de sa qualité ("Il n'est pas nécessaire de mesurer la qualité des données provenant de ce système, nous sommes conscients de sa médiocrité").

Il est important d'introduire le concept de programme continu, pour passer, à maturité, d'un mode réactif, caractérisé par une série de projets ponctuels, à un mode proactif. Cela garantira également que les attentes associées soient réalistes.

- **Mode proactif**

Après avoir établi des bases solides et surmonté les obstacles, le programme de qualité peut se concentrer sur le renforcement des connaissances. Cela implique la formation systématique des nouveaux employés et des sessions de mise à jour régulières pour les anciens. À terme, un plan de développement des compétences adapté à chaque type d'acteur doit être établi.



## 11. Réaliser des REX pour favoriser le passage à l'échelle

Le passage à l'échelle d'un processus de qualité des données (QDD) peut s'avérer être un défi complexe, en particulier pour les grandes organisations qui gèrent des volumes importants de données disparates. Ce défi est souvent amplifié par la nature évolutive des données et des besoins métiers, ce qui implique une adaptation et une amélioration continue du processus QDD.

Il est alors important de réaliser un retour d'expérience (REX) rigoureux afin d'analyser les succès et les leçons apprises lors du déploiement initial du QDM sur un périmètre défini. Cette analyse fournit des informations précieuses pour optimiser le processus et faciliter son passage à l'échelle dans l'ensemble de l'organisation. En d'autres termes, le REX permettra :

- **L'identification des points forts et des faiblesses** : Le REX permet d'identifier les aspects du processus qui ont bien fonctionné et ceux qui nécessitent des améliorations. Cela permet de capitaliser sur les réussites et de corriger les erreurs avant de déployer à plus grande échelle.
- **La compréhension des défis et des risques** : Le REX met en lumière les défis et les risques rencontrés lors du déploiement initial. Cela permet de mieux les anticiper et de mettre en place des stratégies d'atténuation lors du passage à l'échelle.
- **L'amélioration continue du processus** : Le REX favorise une approche d'amélioration continue du processus QDM. En tirant des leçons des expériences passées, il est possible d'optimiser continuellement le processus pour répondre aux besoins évolutifs de l'organisation.
- **La communication et collaboration** : Le REX encourage la communication et la collaboration entre les différentes parties prenantes impliquées dans le déploiement du QDM. Le partage d'expériences et de connaissances favorise une compréhension commune et une approche cohérente pour le passage à l'échelle.
- **La justification des investissements** : Le REX fournit des données concrètes pour justifier les investissements réalisés. En démontrant les avantages et les bénéfices tangibles du processus, il est possible d'obtenir un soutien continu de la part des dirigeants et des parties prenantes.



## Question 4 : Quels sont les facteurs clés de succès d'une démarche de gestion de la qualité des données ?

### ✚ Engager le soutien de la Direction pour établir une culture de la qualité des données pour faciliter :

- L'allocation des ressources nécessaires et du budget pour : remédier aux problèmes, former et faire monter en compétences, acquérir l'outillage adapté
- L'alignement avec la stratégie globale de l'entreprise
- L'établissement des normes et politiques
- La valorisation, promotion, l'encouragement et la reconnaissance des personnes impliquées dans le processus

### ✚ Maîtriser les risques identifiés en amont pour :

- Réduire les coûts
- Satisfaire les clients en évitant les problèmes de qualité
- Se conformer à la réglementation

### ✚ Définir des objectifs clairs, précis et mesurables pour servir de boussole à l'initiative

Les objectifs mesurables permettent de suivre les progrès et d'identifier les domaines où les efforts ne sont pas suffisants. En suivant les KPIs, nous pouvons identifier rapidement les problèmes de non-qualité et prendre les mesures correctives.

### ✚ Sensibiliser et former

La sensibilisation aide à comprendre pourquoi les données sont importantes pour l'organisation. La formation permet d'adopter les bonnes pratiques en matière de collecte, de traitement et de gestion de la qualité des données. Elle contribue à la création d'une culture organisationnelle qui valorise la qualité des données.

### ✚ Suivre les progrès au fil du temps

Une initiative autour de la qualité des données n'est pas un projet avec une date de début et une date de fin, c'est une amélioration continue qui doit être intégrée dans la culture de l'organisation. Cela nécessite un engagement à long terme et une approche proactive comprenant la mise en place de normes, nettoyage et corrections réguliers.

### ✚ Etablir une collaboration cross-métier pour garantir une approche holistique

La qualité des données est transverse et nécessite une gouvernance collaborative pour atteindre les objectifs fixés.

# Merci à tous les Contributeurs au Cahier n°3 de Data Quality

- Régis Deshayes – Zeiss
- François Blondel – Conseil et Perspectives, Architecte d'Entreprise
- Marguerita Kanaan – PwC, Manager Data & IA
- Anzize Badarou – Lintésens, Directeur DATA - BI, Expert en Gouvernance des Données
- Ibtissem Mokhtari – Orange Digital Services, Consultante Senior Data Gouvernance
- Othman Issaad – Orange Business Services
- Mohamed Sylla –Deloitte, Senior Manager Technology & Data Risk
- Kodjani, Kossi Jean Lucien Ganyo –Deloitte, Assistant Manager Technology & Data Risk
- Benoît Piau – Innoval, Directeur Assurance Qualité et Organisation
- Marcel Lee – PwC, Associé Data & IA
- Isabelle Wattiau – Board Dama France