

Groupe de Travail

QUALITÉ DES DONNÉES

cahier
pratique

n°
2

*POINTS DE REPÈRE CLÉS
POUR MESURER LA QUALITÉ
DES DONNÉES.*



ISACA[®]

AFAI | PARIS-FRANCE CHAPTER

Groupe de Travail
QUALITÉ DES DONNÉES

cahier
pratique

n° 2

**POINTS DE REPÈRE CLÉS
POUR RÉALISER LA MESURE
DE LA QUALITÉ DES DONNÉES**

TABLE DES MATIÈRES

INTRODUCTION	p. 4
PRÉSENTATION de l'ISACA-AFAI et de DAMA France	p. 6
Liste des participants au Groupe de Travail DATA QUALITY	p. 9
1/ Comment déterminer les données à contrôler et qui feront l'objet d'un reporting / revue périodique sur la qualité ?	p. 10
2/ Comment prioriser les dimensions (critères qualité) qui vont faire l'objet de mesure et cela a-t-il un sens ?	p. 14
3/ Comment exprimer un besoin qualité / formaliser une règle qualité ?	p. 18
3.1 / Exprimer le besoin	
3.2 / Fixer un seuil de qualité	
4/ Comment déterminer le placement des contrôles sur le cycle de vie de la donnée et comment faire les contrôles ?	p. 22
5/ Comment constituer des indicateurs agrégés de mesure qualité ?	p. 26
6/ Quels outils pour mesurer la qualité et en faire le suivi / reporting ?	p. 32
7/ Ouverture sur nouvelles approches pour analyser la qualité / identifier des anomalies	p. 36

POURQUOI UN GROUPE DE TRAVAIL SUR LA QUALITÉ DES DONNÉES?

Pour la grande majorité des entreprises de taille significative, il ne fait aujourd'hui plus aucun doute que les **données** constituent un **actif majeur**.

Pour autant, de nombreuses initiatives autour des données rencontrent d'importantes difficultés lors de leur mise en œuvre, en particulier en lien avec la qualité des données.

Forts de ce constat, **DAMA France** et **l'ISACA-AFAI** ont recueilli de nombreuses demandes d'entreprises sur la gestion de la qualité des données.

Les deux associations se sont alors **rapprochées** pour **promouvoir et développer les avancées** sur ce sujet en constituant un **Groupe de Travail**.

Le **Groupe de Travail « Data Quality »**, à la composition diversifiée, a pour objet d'éclairer et de **fournir une aide concrète** aux entreprises sur le **sujet de la gestion de la qualité des données**, à travers des **livrables et des illustrations concrètes et pratiques**.

Les restitutions des travaux prendront des formes diverses, notamment des **cahiers pratiques** sur des thématiques ciblées.

Marcel Lee – Président DAMA France
Vincent Manière – Président ISACA-AFAI

POINTS DE REPÈRE CLÉS POUR RÉALISER LA MESURE DE LA QUALITÉ DES DONNÉES

La qualité des données fait écho au degré de **conformité d'une donnée face à ses usages et consommateurs**. Ce degré de conformité se doit d'être objectivé de manière quantitative, en d'autres termes, mesuré.

Après un **1^{er} cahier pratique** introduisant les **Approches tactiques et stratégiques sur la qualité des données**, le **cahier pratique n°2** a pour ambition d'apporter **les points de repère clés pour réaliser la mesure de la qualité des données**, du ciblage au reporting qui en ressort.

Le cahier décrit les **approches considérées comme « standard »** par les praticiens de la gestion de la qualité des données, et ouvre une discussion sur de **nouvelles approches** faisant appel à l'Intelligence Artificielle.

PRÉSENTATION DES ASSOCIATIONS



AFAI | PARIS-FRANCE CHAPTER

L'ISACA-AFAI est l'association de référence pour un digital de confiance.

Créée en 1982, elle est le chapitre français de l'ISACA, association internationale qui regroupe plus de **165 000 membres** dans **188 pays** ; elle rassemble aujourd'hui près d'un millier de professionnels en France.

Sa vocation est d'aider les organisations à mettre le système d'information au service de leur stratégie et de leur développement, à travers trois axes :

- Contribuer à la réflexion, via des groupes de travail et des événements, sur les sujets clés de la gouvernance du numérique et de la sécurité des SI.
- Favoriser l'utilisation en France de référentiels et certifications de l'ISACA.
- Former les professionnels, en les préparant aux certifications professionnelles internationales délivrées par l'ISACA autour de la gouvernance et de la maîtrise des SI (CISA, COBIT, CISM, CRISK, CGEIT, COBIT, CSX...), mais aussi à la gouvernance et au management des SI.

L'ISACA-AFAI contribue ainsi à la création de valeur pour les métiers, à améliorer les compétences et la performance de l'informatique, ainsi qu'à maîtriser les risques liés au système d'information.

L'ISACA-AFAI rassemble une large communauté de professionnels, qui partagent leur engagement pour des systèmes d'information de confiance et créateurs de valeur :

- directeurs des systèmes d'information,
- auditeurs internes et externes,
- consultants,
- directeurs du digital, ou de l'innovation,
- responsables de la Sécurité des Systèmes d'Information (RSSI),
- directeurs financiers ou contrôleurs de gestion,
- risk managers,
- juristes...

Par ailleurs, active pour développer la féminisation des métiers du numérique, l'ISACA-AFAI est membre du collectif Femmes@Numérique.

NOTRE MISSION VOUS AIDER À BÉNÉFICIER DE TOUTES LES POSSIBILITÉS DU NUMÉRIQUE
NOTRE PROMESSE LA CONFIANCE AU SERVICE DE L'INNOVATION TECHNOLOGIQUE

isaca-afai.fr



DAMA (Data Management Association) International est une association mondiale créée en 1980 à but non-lucratif, regroupant des professionnels du Data Management, dont de nombreux auteurs d'ouvrages techniques, conférenciers, et autres référents de l'industrie.

Le but de DAMA International est de promouvoir la compréhension, le développement, l'adoption et la pratique du Data Management au sein des organisations, qu'elles soient publiques, privées, ou associatives.

DAMA International se positionne comme une ressource essentielle pour les professionnels du Data Management et de la Gestion de l'Information, afin de les aider à approfondir leurs connaissances et compétences, notamment via

- la mise à jour et la publication du Data Management Body of Knowledge (DMBOK), ouvrage de référence du domaine,
- le développement d'un programme de certification «Certified Data Management Professional» (CDMP),
- la parution d'études de cas,
- le partage d'expériences et de bonnes pratiques entre professionnels,
- l'organisation de conférences dédiées au Data Management.

Aujourd'hui, l'association regroupe 70 chapitres locaux sur les cinq continents, et compte plus de 20 000 membres à travers le monde.

DAMA France, créé en 2014, est le chapitre français de l'association DAMA International. Comme DAMA International, DAMA France est une association à but non-lucratif, regroupant des professionnels du Data Management en France.

Les objectifs de DAMA France sont de :

- accompagner, favoriser l'acculturation et la montée en compétence des acteurs du Data Management en France,
- innover et participer à l'élaboration et promouvoir les meilleures pratiques du Data Management en France,
- fédérer les initiatives françaises dans le domaine du Data Management,
- créer des réseaux locaux de rencontres de professionnels du Data Management, susciter l'échange entre praticiens.

Aujourd'hui, DAMA France compte plus de 150 adhérents, dont plus de 40 entreprises, et :

- organise de manière trimestrielle un Club des Chief Data Officers des organisations françaises (plus de 30 sociétés membres),
- anime des webinars ou afterworks mensuels sur le Data Management,
- propose du contenu dédié au Data Management sur son site Internet.

dama-france.org



LISTE DES PARTICIPANTS AU GROUPE DE TRAVAIL **DATA QUALITY**

(par ordre alphabétique)

Antoine Mantelier – BPCE

Aude de Touchet – Enedis

Bernard Ruysen – AXA's Corporate Center

Christophe Bonnefoux – BNP Paribas Asset Management

Christophe Hesters – Safran

Florence Rivière – Société Générale

François-Benjamin Remazeilles – Nespresso, VP Organization & Logistics DAMA France

Ghislaine Doukhan – Safran

Hélène Naudon – Groupe La Poste

Isabelle Pires – BNP Paribas Asset Management

Juliette Bechdolf – AXA's Corporate Center

Marcel Lee – Pramana, Président de DAMA France

Matthieu Duponchel – Mazars

Mathieu Minault – Société Générale, VP Marketing DAMA France

Olivier Dardenne – Safran

Olivier Tia Tiong Fat – Société Générale

Philippe Rideau – Société Générale

Samia Belhocine – Groupe La Poste

Sandra Holgado – Enedis

Thibaut Fasquel – Caisse des Dépôts

Valérie Houssin – Caisse des Dépôts

Vincent Givaudan – Pramana, VP Education DAMA France

Vincent Manière – VMA Conseils, Président de l'ISACA-AFAI

Virginie Cuvilliez - RSM France, Secrétaire Générale de l'ISACA-AFAI



1 DONNÉES À CONTRÔLER

COMMENT DÉTERMINER LES DONNÉES À CONTRÔLER ET QUI FERONT L'OBJET D'UN REPORTING / REVUE PÉRIODIQUE SUR LA QUALITÉ ?

Dans un monde idéal, il serait intéressant pour une organisation d'avoir à disposition un reporting sur la qualité de l'ensemble de ses données afin de pouvoir toutes les surveiller et de disposer, quelle que soit la situation, de toutes les informations nécessaires. Dans la réalité, au vu de l'ampleur de la tâche, il est impossible de superviser toutes les données d'une organisation de façon pertinente et effective.

Il faut donc **faire des choix et cibler les données** que l'on souhaite mettre sous contrôle sous peine de diluer les efforts investis, de perdre en efficacité et in fine de ne pas aboutir à des résultats probants. Le cas échéant, et non souhaité, cela reviendrait à gaspiller les ressources allouées à la qualité des données qui sont de plus très souvent difficiles à mobiliser comme expliqué dans le 1^{er} cahier pratique.

Plusieurs questions vont ainsi se poser : combien de données mettre sous contrôle ? Lesquelles sélectionner ? Pourquoi telle donnée plutôt que telle autre ? etc. ... Il n'existe malheureusement **pas de réponse unique** à ces questions et chaque organisation devra trouver ses propres vérités.

Pour ce faire, la première chose à réaliser est de clarifier les objectifs recherchés par l'organisation au travers de la qualité des données. S'agit-il de limiter des risques réglementaires ou financiers ? S'agit-il plutôt de gagner en efficacité opérationnelle ou de permettre de nouveaux usages ? Ou bien encore est-ce une question d'image et de réputation ? Les réponses à ces questions vont non seulement influencer le choix des données à mettre sous contrôle mais également la méthode pour les identifier.

En effet, il n'existe pas une unique méthode pour prioriser les données à mettre sous contrôle mais plusieurs qui seront plus ou moins pertinentes en fonction du contexte de l'entreprise, de sa maturité sur les données et des objectifs que l'organisation cherche à atteindre via la qualité des données.

Afin de synthétiser les choses, on peut identifier **deux grands types d'approches qui se distinguent** : les **approches par la valeur des données** et les **approches par les produits finis**.

Les **approches par la valeur des données** identifient et priorisent les données en fonction de leur importance intrinsèque au sein de l'organisation, mettant sous contrôle les données ayant le plus de valeur pour l'organisation. La **valeur** n'est pas vue ici comme une somme financière précise, qui est de surcroît très difficilement quantifiable, mais est **estimée au travers de critères de priorisation** qui vont permettre de dégager les données à superviser.

Ceux-ci sont à déterminer en fonction du contexte de l'organisation ; on peut citer, parmi les plus fréquemment retenus, la transversalité des données : plus une donnée est transverse aux processus de l'entreprise ou utilisée par un grand nombre de consommateurs, plus elle a de valeur. Par exemple, les données de références répondent parfaitement à ce critère.

Un autre critère très souvent utilisé est la criticité de la donnée pour le ou les métier(s). Elle vise à identifier les données de première importance pour l'accomplissement de la stratégie métier de l'organisation et qui soutiennent les processus critiques de l'entreprise. Par exemple, on va retrouver des données réglementaires, des données de logistiques ou encore des données visibles par les clients. Pour réaliser cette sélection, ces approches nécessitent une bonne connaissance des données et de leurs incidences sur les processus de l'entreprise. Elles s'appuient généralement sur un glossaire métier ou sur des travaux de catalogage des données à mener au préalable (dictionnaire de données).

● DONNÉES CATÉGORISÉES PAR DOMAINE

- Type de contribution: Entrant, Sortant ou Entrant/Sortant
- Niveau de contribution
- Niveau de satisfaction actuel

● MISSIONS

- Priorité de chaque mission

	MISSIONS DE L'AGENCE	PRIORITÉ DE LA MISSION	DOMAINE DE DONNÉES							
			Donnée A			Donnée B				
			Type de contribution	Niveau de contribution	Niveau de satisfaction actuel	Type de contribution	Niveau de contribution	Niveau de satisfaction actuel		
Domaine de mission	Mission 1									
	Mission 2									
	Mission 3									

Exemple de cadre d'alignement entre les missions d'une entreprise et les données (approche par la valeur des données)

¹ Le glossaire métier décrit de façon transverse les connaissances sur les données à un niveau métier : définition, entité propriétaire, taxonomie, règle et exigences métiers, lien entre les données (modèle d'information)...

Le glossaire métier sert aux utilisateurs métier, Data steward et à certains utilisateurs techniques (architecte, custodian,...)

² Le dictionnaire de données décrit les connaissances sur les données dans chaque système à un niveau technique : définition, structure, modèle de données, liste de valeurs, exigences de performances d'accès,...

Le dictionnaire de données sert aux Data steward et aux utilisateurs techniques (architecte, custodian,...)

Plutôt que de se focaliser sur les données et leur valeur intrinsèque, les **approches par les produits finis** se concentrent dans un premier temps sur les finalités qui sont produites par les processus de l'organisation : rapports, bilan comptable, publications, ... L'objectif est d'**identifier et prioriser les finalités des produits finis à mettre sous contrôle en raison de leur caractère essentiel ou critique**, c'est-à-dire celles pour lesquelles la non qualité (ou disqualité) aura le plus d'impact sur l'organisation. Par exemple, ce peut être des reportings réglementaires, des indicateurs de pilotage clés remontés au COMEX, etc.

Une fois cette priorisation effectuée, on identifie, pour chaque finalité critique, **les données qui participent à leur construction depuis leur origine jusqu'à la destination avec les transformations qu'elles subissent sur ce tracé (« lineage de données »)**. Cela permet de connaître quelles données il est utile de mettre sous contrôle pour l'organisation et comment leur qualité participe à la qualité finale des produits finis qui peut elle-même être ainsi évaluée.

Ces deux approches ne sont bien entendu **pas exclusives** et peuvent très bien être **hybridées** d'autant qu'elles ont chacune leurs intérêts et leurs risques. Par exemple, les approches par la valeur des données sont censées maximiser le retour sur investissement au global pour l'organisation étant donné qu'elles ciblent les données à plus forte valeur. La problématique est qu'il est souvent très difficile de prouver la valeur des résultats apportés car ils sont très diffus et non totalement connus. Cela induit une complexité dans la promotion des actions de qualité et dans la démonstration du retour sur investissement. A contrario, les approches par produits finis permettent de plus facilement identifier les gains attendus et réalisés par les actions de qualité, mais génèrent en retour moins de bénéfices induits ou indirects au global pour l'organisation.

Tableau de synthèse des éléments relatifs aux deux types d'approches

	APPROCHE PAR LA VALEUR DES DONNÉES	APPROCHE PAR LES PRODUITS FINIS
OBJECTIFS / FOCUS	<ul style="list-style-type: none"> Gain d'efficacité, d'image et de réputation Promotion des nouveaux usages des données 	<ul style="list-style-type: none"> Objectifs réglementaires Usages métiers ciblés par la stratégie métiers Amélioration du pilotage de l'organisation
MATURITÉ	<ul style="list-style-type: none"> Nécessite une maturité sur la connaissance des données, et une maturité des décideurs. Nécessite d'identifier des cas d'usage porteurs de valeur 	<ul style="list-style-type: none"> Nécessite de remonter la chaîne de production des produits finis Nécessite de faire collaborer de nombreux acteurs métier et SI
AVANTAGES	<ul style="list-style-type: none"> Bénéfices plus systémiques Maximisation du retour sur investissement 	<ul style="list-style-type: none"> Résultat visible et sur le court terme
INCONVÉNIENTS	<ul style="list-style-type: none"> Valeur difficile à démontrer, qui plus est, à court terme Nécessité que les décideurs soient a priori convaincus des bénéfices 	<ul style="list-style-type: none"> Moins d'effet de levier



PRIORISER LES CRITÈRES QUALITÉ

2

COMMENT PRIORISER LES DIMENSIONS (CRITÈRES QUALITÉ) QUI VONT FAIRE L'OBJET DE MESURE ET CELA A-T-IL UN SENS ?

Nota Bene : Dans l'ensemble du document, nous utilisons le terme « dimension » à la place de « critère » qualité pour évoquer une caractéristique mesurable de la donnée, en conformité avec la terminologie DAMA-DMBOK 2. Ce terme « dimension » n'est donc ici aucunement lié à la notion de « dimension » ou « axe d'analyse » des modèles dimensionnels qu'on peut voir dans le domaine de la Business Intelligence.

Tout d'abord, rappelons rapidement ce que sont les dimensions de qualité. Une dimension de qualité des données est une **caractéristique ou un élément mesurable de la donnée**. Les dimensions de qualité de la donnée permettent de catégoriser les critères de qualité sur les données afin d'en :

- Faciliter la compréhension des critères ;
- Permettre de fixer un **vocabulaire** commun dans l'organisation ;
- Disposer de notions communes de référence pour identifier les contrôles ;
- Faciliter la **gestion** des critères et des **contrôles** de qualité, par exemple si on veut constituer une bibliothèque.

Il est donc nécessaire de s'être mis d'accord sur le choix de ces dimensions pour l'organisation, sur leur définition afin qu'elles soient partagées, comprises et utilisables. Toutefois malgré ce fondamental, le constat est sans appel : **toutes les dimensions ne se valent pas dans leur facilité de compréhension et de mise en œuvre**. Par exemple, la complétude parlera au plus grand nombre là où la « cohérence » fera plus volontiers débat.

Mais revenons à notre question. La tournure de la question s'applique dès lors qu'en tant que praticien de la qualité des données, on cherche à réaliser un travail transverse, d'agrégation pour anticiper et traiter les problèmes de qualité au global.

Or les données sont corrélées à l'usage qui en est fait d'une part – en effet, pour rappel, la qualité des données se réfère à la conformité de la donnée par rapport à l'usage qu'on veut en faire. D'autre part, lorsqu'on regarde de plus près, c'est tout un « cycle de vie de la donnée » qui émerge et qu'il faut éventuellement étudier, du moins sur la partie du cycle qui mène à cet usage à mettre sous contrôle. Par exemple le besoin de complétude dans un processus d'acquisition client n'est pas le même que dans un processus de contractualisation ou de facturation.

Ainsi le **choix des dimensions se fera en fonction de l'usage de la donnée, de son cycle de vie, et de la facilité de mise en œuvre**, qui sont directement dépendants des processus métiers et du système d'information.

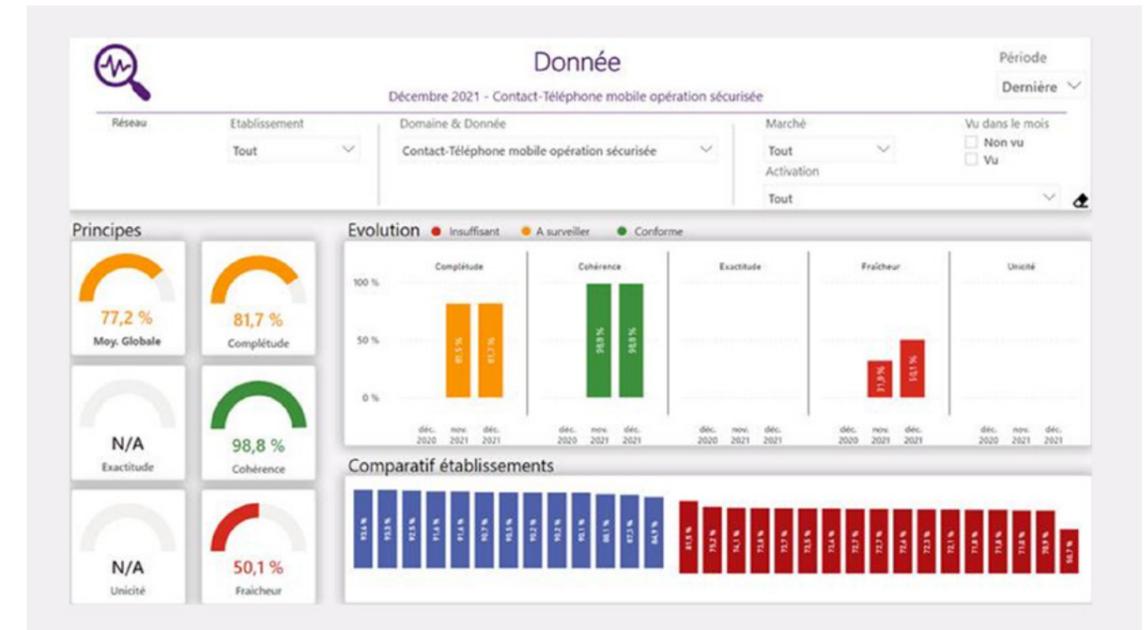
³ Le cycle de vie d'une donnée est l'ensemble des étapes par lesquelles une donnée va passer depuis sa collecte / création jusqu'à sa suppression, en passant par sa transformation, son stockage, ses usages, etc.

Sans être exhaustif, les dimensions fréquemment rencontrées sont répertoriées dans le tableau suivant :

DIMENSION	DESCRIPTION	EXEMPLE DE DISQUALITÉ ET IMPACT	COMMENT DÉTECTER LA DISQUALITÉ
Complétude	Donnée renseignée, pas de donnée manquante	Date de naissance de clients non renseignée Impact : Campagne marketing moins pertinente car la segmentation des clients selon leur âge n'est pas possible	- Vérifier qu'il n'y a pas de valeurs manquantes (champs vides, N/A, #ERREUR, ...) - Vérifier qu'une partie de données n'est pas manquante en comparant avec un back-up ou avec les métadonnées en vérifiant l'heure de création pour connaître la cause racine de la disqualité : Erreur humaine ou bug informatique de remontée de donnée, etc.
Intégrité	Donnée non corrompue ou modifiée sur son cycle de vie	Bilan financier certifié modifié après publication Impact juridique fort	- Comparer une donnée avec une copie de référence - Analyser le cycle de vie en traçant les modifications réalisées sur la donnée
Fraicheur	Données à jour ou disponible au moment opportun	Non réception de la facture par un Client après mise à jour de son mail en ligne (lenteur du délai de mise à jour des données) Impact financier pour l'entreprise : à cause de retard dans le paiement des factures par le client	- En interrogeant les SI sur la fraîcheur de la mise à jour de l'application - En regardant l'horodatage de la collecte de la donnée
Unicité	Donnée non redondée dans une source	Règlement en double d'une facture à cause de doublons dans la base de données fournisseurs Impact financier pour l'entreprise	- L'unicité se vérifie au sein d'une même source et se détecte en cherchant les valeurs en double, soit sur une colonne unique soit sur un ensemble de colonnes.
Exactitude	Donnée correspondant à la réalité qu'elle décrit	Une base de données avec des adresses mails d'agents de l'entreprise à la place des adresses mails des fournisseurs. Impact d'image: des fournisseurs non satisfaits à cause retard dans le règlement de leurs factures	- Vérifier que les valeurs des données sont valides et correctes, c'est-à-dire, qu'elles se trouvent dans une fourchette valide - Faire appel à des techniques de visualisation de données (ex. nuage de points) pour repérer les valeurs invalides et aberrantes.
Vraisemblance	Donnée plausible, dans un éventail de valeurs crédibles ou ayant du sens dans son contexte d'utilisation	Un assuré né en 1900 Impact : donnée non exploitable	- Contrôle par rapport à une liste de valeurs - Contrôle de format, etc
Cohérence	Valeur de la donnée non divergente avec la valeur d'autres données avec lesquelles elle est corrélée	Une personne est à la fois dans la base de donnée des prospects et dans la base de donnée client. Cette incohérence peut être à l'origine d'envoi de relances non nécessaires Impact d'image	- Réconcilier les données et les comparer avec une autre source pour mettre en lumière les incohérences - Analyser la cohérence par rapport à d'autres sources qui portent la même information ou d'autres données qui sont corrélées à cette donnée
Validité	Donnée appartenant à une liste de valeurs prédéfinies	Un assuré avec un type différent de la liste valeur ('Particulier', 'Entreprise') Impact : donnée non exploitable	- Contrôle par rapport à une liste de valeurs - Contrôle de format, etc

En fonction de la maturité de l'entreprise, le « set » de dimensions que l'on adoptera pourra être différent. Lorsque l'entreprise débute une initiative qualité, elle voudra en général se doter d'un « set » minimum de dimensions systématiques et standards à mesurer, car cela aide à se dire qu'on commence à mesurer et objectiver la qualité. Le risque est de ne pas suffisamment sensibiliser les décideurs (et les métiers) aux corrélations existantes entre usages et métriques. Il faudra donc absolument communiquer sur les limites de l'exercice.

Avec la montée en maturité sur la qualité des données, on aura à la fois des « sets » d'indicateurs standardisés et systématiques (potentiellement plus étendus, mais pas nécessairement), et surtout des choix de dimensions plus corrélés à l'usage. On sera alors dans un pilotage plus fin, et davantage associé à un objectif opérationnel.



Exemple de dashboard Data Quality avec des dimensions pertinentes / non pertinentes



Exemple de dashboard Data Quality – cohérence entre deux dates liées

FORMALISER UNE RÈGLE QUALITÉ

3

COMMENT EXPRIMER UN BESOIN QUALITÉ / FORMALISER UNE RÈGLE QUALITÉ ?

3.1 Exprimer le besoin

Le sujet des dimensions nous indique 3 fondamentaux pour traiter les sujets de qualité d'une donnée : l'usage, le cycle de vie et la facilité de mise en œuvre.

Ainsi **on exprime un besoin qualité au regard de l'usage et des problèmes de qualité** souvent bien connus des métiers ou de la DSI (Data Owner, Data Steward, Data Custodian). Ce besoin s'exprimera d'abord par des règles qualité, qui expriment « l'état dans lequel la donnée doit se trouver pour être utilisable ».

Les dimensions apportent une aide dans l'expression de ces règles qualité pour s'assurer que rien n'a été oublié. Pour ceux qui se posent la question Règle / Dimension, voici ce que nous proposons :

- Une règle s'applique sur une seule dimension
- On peut avoir plusieurs règles pour une dimension

Le besoin de qualité évolue suivant le cycle de vie de la donnée : on peut avoir une même règle contrôlée plusieurs fois sur différents endroits du cycle de vie de la donnée avec des paramètres différents (seuils etc...)

Cycle de vie d'un projet d'investissement : de la création jusqu'à la mise en place



Exemple de cycle de vie d'un projet d'investissement, avec les indications de Création [C], Modification [U] sur les données le long du cycle de vie

3.2 Fixer un seuil de qualité

Lorsque l'on détermine une règle, on en ressort a priori **un indicateur s'appuyant sur un dispositif d'évaluation** (contrôles, profilage des données...). Cet indicateur va permettre de prendre une décision ou d'enclencher une action. Il est d'usage, pour faciliter la prise de décision de fixer un seuil, à savoir une valeur en dessous de laquelle la qualité est considérée comme non admissible par rapport à l'usage qu'on veut en faire.

Ce seuil est directement corrélé à un impact et donne du sens à l'indicateur pour la prise de décision ou d'action.

Ex : une base clients permet de bien connaître les clients pour leur proposer des services ciblés ; si des doublons sont présents, on observera une perte de temps en raison des retraitements ; l'impact est ici opérationnel et financier. Le juste niveau de qualité est le niveau en dessous duquel la perte de temps sera jugée non admissible pour l'entreprise.

Le seuil étant lié à un impact, sa fixation fera intervenir une évaluation d'impact (que ce soit opérationnel, financier, etc.), mais aussi l'effort à réaliser pour être au-dessus. Plus le seuil est élevé – c'est-à-dire exigeant – plus l'effort et les coûts induits seront importants – essentiellement parce que des moyens substantiels de mise sous contrôle et d'automatisation devront certainement être mobilisés.

Focus : faut-il déterminer le seuil a priori lors de la définition de la règle qualité ?

Lorsque l'on détermine une règle qualité, il n'est pas toujours aisé de fixer immédiatement un seuil, et encore moins le « bon » seuil, sauf à disposer d'un minimum de connaissance et d'expérience à la fois sur les usages métier que l'on adresse et d'avoir une bonne compréhension du rapport impact vs. effort lié à ce seuil.

Ainsi, selon le niveau de maturité de l'entreprise, il sera possible d'accepter dans un premier temps que le seuil ne soit pas fixé du tout, ou qu'il soit fixé de manière indicative et arbitraire. Le seuil sera alors revu au fil du temps selon les efforts de remédiation subséquents et les progrès accomplis.

Dans le cas où il est fixé de manière indicative, sa valeur peut être définie sur celle de la première observation de mesure réalisée. On pourra y ajouter un « delta » dépendant des efforts à entreprendre pour avoir, in fine, un impact métier acceptable.

La fixation du seuil est donc un exercice empirique et itératif de manière générale, et faisant appel à la fois à des compétences métier (pour l'évaluation de l'impact admissible et des efforts métier) et des compétences SI (en particulier pour l'évaluation des efforts de redressement).

Enfin, l'expérience montre qu'il est nécessaire de bien connaître la donnée avant de fixer un seuil pertinent et que cela peut passer par une analyse (stratification, répartition) d'échantillons largement représentatifs de cette donnée, permettant de s'assurer de sa cohérence notamment lorsque plusieurs sources sont à son origine.

Le seuil : un chemin ou l'objectif ultime ?

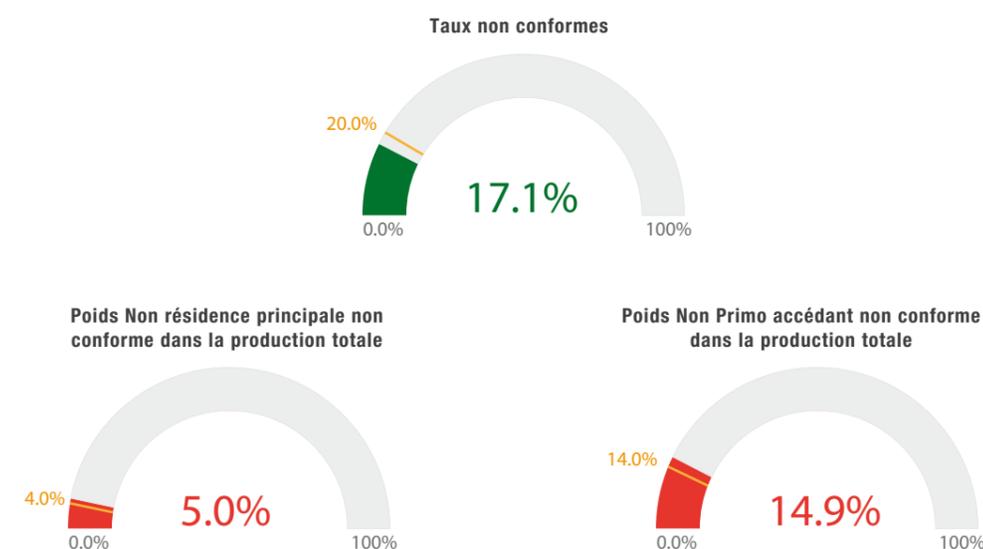
Nous avons évoqué le seuil comme le déclencheur pour les actions de remédiation à mener et d'effort à porter sur les données. Le seuil est lié au périmètre, au moyen, et à l'objectif visé. **Ainsi ce seuil peut être vu de deux manières : soit comme l'étape d'un chemin soit comme un l'objectif ultime à atteindre.**

Lorsque l'on initie des travaux sur la qualité de la donnée, on aura plutôt tendance à se fixer une première étape réaliste afin qu'un premier palier puisse être atteint. Il s'agit via ce premier objectif atteignable d'encourager les parties prenantes dans les efforts qu'elles vont devoir mettre en œuvre dans un contexte métier ou SI le plus souvent déjà sous tension.

Avec la progression des pratiques et de la maturité globale de l'organisation, il sera ensuite possible de mettre en place à la fois un **seuil de référence**, en dessous duquel des actions de remédiations immédiates devront être réalisées pour revenir au seuil de référence (« data quality operations », qui peuvent être agrémentées de SLA – Agréments sur les Niveaux de Service), et un objectif, une **cible idéale de qualité**.

Avec ces éléments en tête, il est ainsi possible d'avoir **un dispositif pour les remédiations court terme et un plan de progrès pour atteindre la cible**, tenant compte de l'ensemble des critères évoqués (usage, impact, effort), tout en gardant en tête que l'effort n'est pas linéaire et que respecter le Principe de Pareto est fortement souhaitable.

Exemple de mesures qualité avec mise en évidence de seuils



⁴Le principe de Pareto, aussi appelé Loi des 80-20, statue qu'environ 80 % des effets sont le produit de seulement 20 % des causes.



CONTRÔLES SUR LE CYCLE DE VIE DE LA DONNÉE

4

COMMENT DÉTERMINER LE PLACEMENT DES CONTRÔLES SUR LE CYCLE DE VIE DE LA DONNÉE ET COMMENT FAIRE LES CONTRÔLES ?

Il y a à minima deux cas à considérer : les contrôles dans un objectif d'assurance qualité pour l'usage et les contrôles dans un objectif de mise sous contrôle globale (souvent lié à un objectif réglementaire ou de gestion des risques).

- Dans un **objectif d'assurance qualité pour l'usage**, le contrôle permet de garantir que la qualité des données est conforme à l'usage qu'on veut en faire. Dans ce cas particulier, il doit donc être **réalisé au plus proche de l'usage**, et non à des moments plus en amont du cycle de vie de la donnée, car la donnée peut subir une dégradation de sa qualité le long de son cycle de vie.

- Dans un **objectif de mise sous contrôle globale**, les contrôles servent à **gérer les risques le long du cycle de vie de la donnée**. Dans cette perspective, il vaut mieux prévoir des contrôles le long du cycle de vie. Mais cela a un coût ; par ailleurs, il n'est pas toujours possible d'agir sur toute la chaîne pour de multiples raisons :

- Organisationnelles (il peut être facile d'agir sur son périmètre mais pas toujours sur le périmètre de l'entité qui nous fournit des données – ceci est particulièrement vrai lorsque l'on travaille dans un écosystème étendu) ;
- Métier : trop de contraintes dans les processus, procédures, et conduite du changement trop chronophage
- Techniques : systèmes trop anciens, investissements trop importants par rapport au gain escompté

Face à ces difficultés, nous considérerons que :

- Lorsque l'**organisation est encore peu mature**, il faut dans un premier temps avoir une assurance qualité pour l'usage qu'on veut faire des données, et donc a minima prévoir un **contrôle au plus proche de l'usage** ;

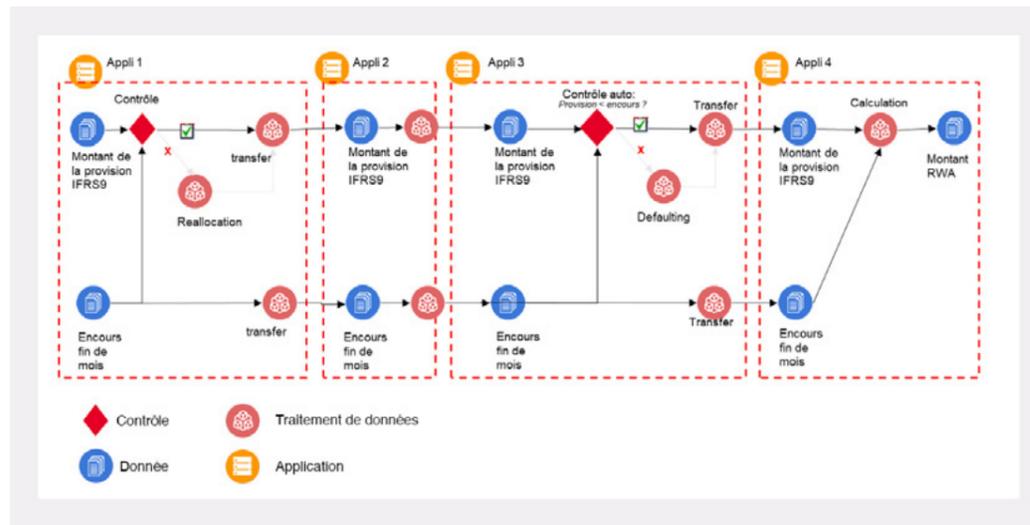
- Avec la maturité grandissante de l'organisation, les points de contrôle vont glisser vers les points de production et de collecte de la donnée (donc aux sources) car c'est le seul moyen viable et pérenne pour limiter le nombre de contrôles, éviter les comportements inappropriés par les consommateurs (ex : modification manuelle de la donnée juste avant son utilisation ou utilisation d'une donnée sans vérifier la pertinence de la source), et acculturer ainsi que responsabiliser les producteurs de donnée.

Focus : le coût des contrôles

Quelle que soit la situation, il est indispensable de reconnaître que tout contrôle a un coût. Il est préconisé de réaliser une étude gain vs. effort avant sa mise en œuvre. Dans un domaine tout autre que la donnée, nous pouvons illustrer ce propos avec l'exemple de la chaîne de prêt-à-porter Primark, qui a décidé de ne plus mettre sur les vêtements les antivol (qui sont des éléments de contrôle), car leur gestion au quotidien (approvisionnement, rattachement et détachement des vêtements par le personnel, anomalies, etc.) avait un coût supérieur au risque avéré de vol de marchandise.

Dans certains cas, l'étude de l'architecture des données, en lien avec le lineage des données, va servir à déterminer des points de contrôle communs et qui permettent par conséquent de réduire les coûts de mise en place. C'est le cas par exemple des contrôles réalisés sur des « hubs » (plateformes) de données.

Exemple de cycle de vie / lineage de donnée avec positionnement des contrôles qualité



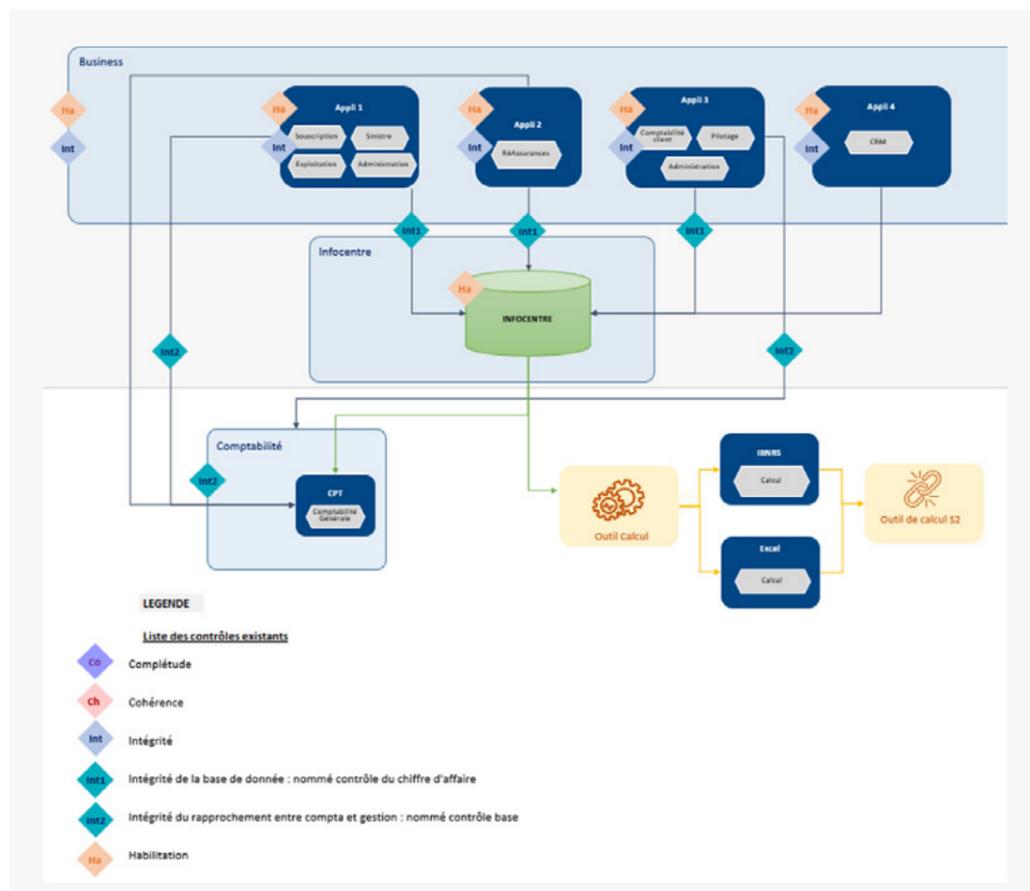
Exemples de règles métier à contrôler :

- [Fraîcheur] Le délai d'approvisionnement d'un article doit avoir été révisé depuis moins de 6 mois ; l'objectif métier est de s'assurer qu'il est toujours « approvisionnable »
- [Complétude] Un article qui va être reçu dans le mois doit disposer d'un emplacement de stockage défini
- [Cohérence] Le nombre de machines déclaré dans le progiciel de gestion doit être égal au nombre de machines déclaré dans le fichier de référence usines.

Exemples de règles qualité (contrôles de cohérence, vraisemblance et validité)

AD01	Revue de l'utilisation des champs dans la calibration	Vérifier que les informations provenant de l'infocentre et utilisées au niveau de la calibration sont bien topées 'critiques'
AD01	Revue de la plage de valeurs adhérents	Vérifier que l'identifiant de l'adhérent utilisé lors de la calibration ne prend pas les valeurs suivantes '0' ou ''
AD03	Revue de la répartition des sinistres par type de défaut	Vérifier que pour le périmètre «Produit-X», les types de statut soient PRODUCT-DEFAULT' ou 'INSOLVENCY'
AD04	Revue de la plage des valeurs des montants de déclaration	Vérifier que le montant en euro des impayés déclarés lors de la déclaration soit toujours supérieur à zéro
AD05	Revue de la cohérence entre la date de déclaration et la date de défaut	Vérifier l'antériorité entre les dates de déclaration et de défaut en fonction du type de risque commercial qui a induit le sinistre

Exemple de cartographie des flux de données et de contrôle qualité existants





5 INDICATEURS AGRÉGÉS DE MESURE QUALITÉ

COMMENT CONSTITUER DES INDICATEURS AGRÉGÉS DE MESURE QUALITÉ ?

Dès lors que les contrôles voient le jour, il devient rapidement souhaitable de les agréger pour disposer d'une vue synthétique de la qualité sur un cas d'usage, un domaine, ou pour toute l'entreprise, et ce, d'autant plus que, comme nous l'avons vu, une mesure qualité est granulaire par nature : par donnée (enregistrement ou dataset), par règle qualité, et par point de contrôle sur le cycle de vie de la donnée.

La constitution de cette vue synthétique va, si elle est bien réalisée, répondre à plusieurs enjeux, potentiellement concomitants :

- Pilotage des travaux qualité ;
- Evaluation du progrès sur la qualité des données ;
- Justification de l'investissement réalisé ;
- Détection de problèmes qualité récurrents ou naissants ;
- Mise sous contrôle dans le temps de la qualité.

Seulement, pour disposer de cette vue synthétique, il est nécessaire d'agréger, et donc potentiellement de décontextualiser les mesures qualité, avec le risque d'avoir alors des indicateurs qui perdent de leur sens. Comment s'y prendre dans ce cas, et existe-t-il d'ailleurs une bonne manière de s'y prendre ?

Les échanges lors du Groupe de Travail convergent tous vers cette conclusion. Il n'y a pas UNE unique bonne manière de s'y prendre. Nous proposons donc plutôt ici des points de repère auxquels il conviendra de se référer si l'on veut agréger des mesures qualité :

1. Bien se mettre d'accord sur la finalité de l'indicateur agrégé

Il n'existe pas un indicateur agrégé qui va pouvoir répondre à toutes les finalités possibles. La cible visée, ainsi que la finalité sont clés pour comprendre les choix qui devront être faits par la suite. Ainsi par exemple, un indicateur agrégé constitué selon une finalité de pilotage global de la démarche qualité sera différent d'un indicateur constitué selon une finalité plus opérationnelle de remédiation et mise sous contrôle de la qualité.

Potentiellement, il sera possible de disposer de plusieurs tableaux de bord qualité, à destination de populations différentes. On veillera dans ce cas à bien communiquer et expliquer le contexte d'utilisation de ces tableaux de bord.

2. Bien spécifier l'indicateur agrégé

Un indicateur agrégé est une donnée, voire même un « produit data » selon la manière dont on voit les choses.

Dans tous les cas, il ne sera utilisable que si les métadonnées métier et techniques qui permettent de le comprendre, de l'utiliser et de le gérer avec une certaine assurance sont bien disponibles et de qualité.

On s'assurera donc à définir l'indicateur, ses règles métier permettant de le calculer, son lineage, la ou les finalités associées, et les précautions d'usage.

3. Un mode de calcul dépendant de la finalité

Le mode de calcul adopté pour un indicateur devra être standardisé, et sera dépendant de sa finalité.

Ainsi, pour une donnée, sur un point de contrôle établi, nous avons potentiellement plusieurs mesures (car plusieurs règles qualité sur différentes dimensions). Une entreprise aura le choix en général entre :

- Prendre la « pire » des mesures et l'accoler à l'indicateur
- Prendre la moyenne arithmétique des mesures sans pondération
- Prendre la moyenne arithmétique des mesures avec pondération

La première option sera à privilégier pour une finalité de remédiation par les opérationnels de la qualité des données afin de détecter une anomalie, et lancer une alerte avec une remédiation potentielle immédiate, tandis que les deux options suivantes seront à privilégier dans une optique de pilotage.

Un mot sur la pondération : les différentes règles qualité n'ont pas toute la même importance, et les seuils respectifs associés peuvent également différer, ce qui milite en faveur d'une pondération par règle qualité. Si celle-ci semble « logique », elle crée cependant une difficulté à la fois dans la compréhension, la lisibilité, et la mise en œuvre comme la maintenance à terme de l'indicateur. Il s'agira donc d'utiliser cette technique avec parcimonie, et de manière homogène. La simplicité est parfois à privilégier en faveur de l'efficacité et de l'efficience.

Une fois cet indicateur établi par donnée, il faudra le remonter dans un tableau de bord. Il pourra être possible d'agréger sur plusieurs données, avec les mêmes considérations que ci-dessus (cf. les 3 options), avec le risque de détériorer davantage la lisibilité et la véracité des indicateurs. En plus des indicateurs purement de qualité, le tableau de bord, selon sa finalité, pourra présenter d'autres métriques comme le taux de couverture de mesure qualité, le nombre d'anomalies ou d'incidents qualité, le niveau de service sur les opérations qualité, le nombre d'itérations pour atteindre un niveau de qualité attendu, etc...

Bien entendu, tous ces éléments seront historisés de manière à pouvoir suivre la situation dans le temps. Si des règles de calcul sont amenées à changer, il faudra bien veiller soit à faire des « pro forma », soit à bien expliciter le changement dans les chiffres.

4. Bien expliciter les tableaux de bord et indicateurs

Chaque tableau de bord et indicateur aura une finalité, mais si c'est une chose de le définir, c'est autre chose d'avoir une bonne compréhension et prise de conscience sur ces éléments par le plus grand nombre, en particulier par les parties prenantes et décideurs concernés.

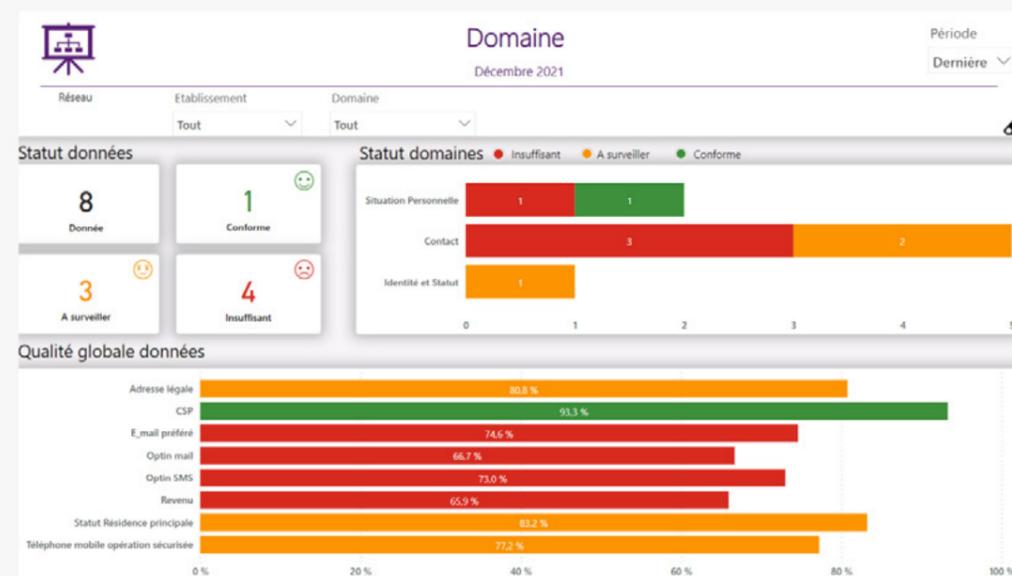
Il faudra donc ne pas oublier de bien expliquer et communiquer systématiquement la portée et les limites de ces tableaux de bord et indicateurs.

Dans tous les cas, on montrera un niveau de qualité au regard d'un ou plusieurs usages, en fonction du mode de contrôle adopté (proche d'un usage unitaire ou pour une multiplicité d'usages). En effet, le risque est de montrer des chiffres qui ne correspondent pas à une situation d'usage, or la qualité, rappelons-le, est une notion contextuelle par nature : il n'existe pas de niveau de qualité « absolu ». Les usages étant amené à changer dans le temps, là encore, il faudra potentiellement prévoir d'avoir des valeurs d'indicateurs qui vont changer dans le temps.

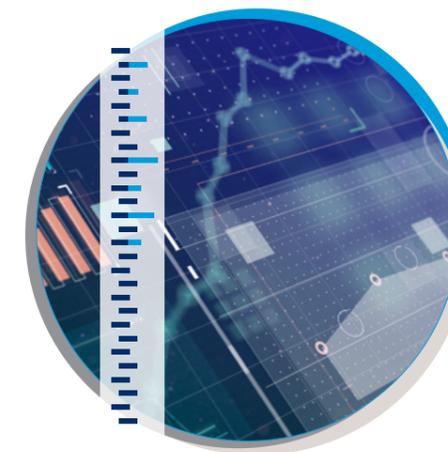
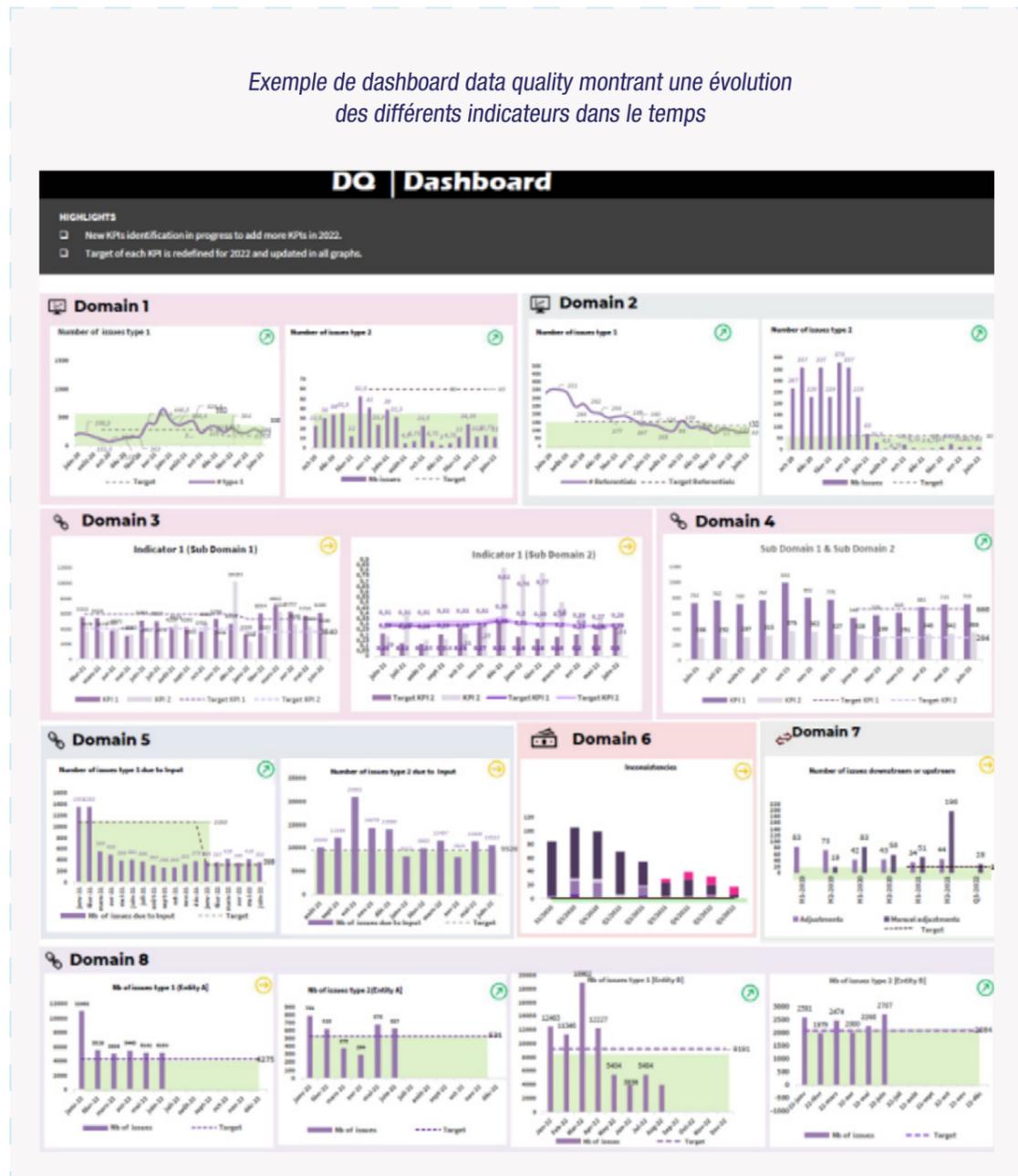
Potentiellement, nous pouvons aussi recommander d'avoir des tableaux de bord qui contextualisent les niveaux de qualité observés avec d'autres indicateurs métier, afin de voir les liens de corrélation et d'influence entre un niveau de qualité et d'autres inducteurs métier (ex : voir un taux de complétude des adresses email et le montant de lead potentiel que représentent les prospects non adressés).

• EXEMPLE 1 DE DASHBOARD DATA QUALITY

Exemple de dashboard sur un domaine, avec une mesure agrégée par donnée, et des indicateurs de conformité (seuils)



Exemple de dashboard data quality montrant une évolution des différents indicateurs dans le temps





OUTILS POUR MESURER LA QUALITÉ ET EN FAIRE LE SUIVI

6

6. QUELS OUTILS POUR MESURER LA QUALITÉ ET EN FAIRE LE SUIVI / REPORTING ?

La gestion de la qualité des données étant une activité revêtant un aspect technique, il est légitime de se poser la question de l'outillage de la mesure de la qualité, et au-delà de l'outillage de la gestion au global des données.

La plupart du temps, pour répondre à cette question, les entreprises pensent à une unique solution « miracle ». Or, elles devraient en premier lieu réfléchir aux « fonctions » de gestion de qualité des données dont elles ont besoin, et voir dans quelle mesure ces « fonctions » sont ou peuvent être couvertes par tout ou partie du système d'information existant. Parmi les « fonctions », on peut en particulier mentionner :

- Profilage de la donnée
- Gestion des règles qualité
- Contrôle des règles qualité
- Reporting opérationnel des mesures qualité
- Alerting
- Assignment et prise en charge des problèmes qualité
- Remédiation des problèmes qualité
- Monitoring continu de la qualité
- Reporting stratégique sur la qualité

L'objectif est de combler l'écart « fonctionnel » dont l'entreprise a besoin, et d'assurer une couverture par différents systèmes.

Exemple de profilage sur des revenus, montrant que les opérationnels ont saisi tantôt des revenus mensuels, tantôt annuels (mise en évidence d'un problème de cohérence)



Au-delà de l'aspect purement « fonctionnel » de la démarche, l'entreprise devra aussi et avant tout s'assurer que l'outillage est compatible avec un état de modèle opérationnel à l'instant t.

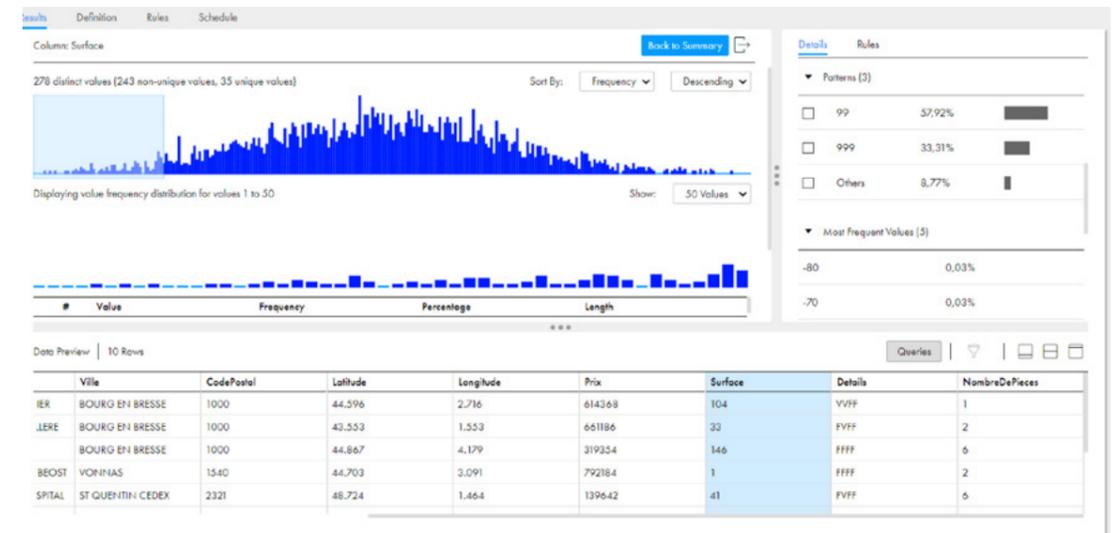
Ainsi, par exemple, certains outils sont difficiles à prendre en main par des opérationnels métier mais si le modèle opérationnel prévoit une prise en charge directement par ces mêmes opérationnels métier de manière décentralisée, il y a risque de non-réalisation des activités en autonomie et donc de perte de temps et d'information — s'il faut effectivement au final prévoir une équipe « technique » en central en sus des opérationnels métier.

Le modèle opérationnel et la maturité de la gestion de la qualité des données étant tous deux amenés à évoluer au fil du temps, il faudra de manière périodique s'assurer que l'outillage mis en place est bien adapté à ces évolutions et le cas échéant, le remanier à la marge, voire intégralement.

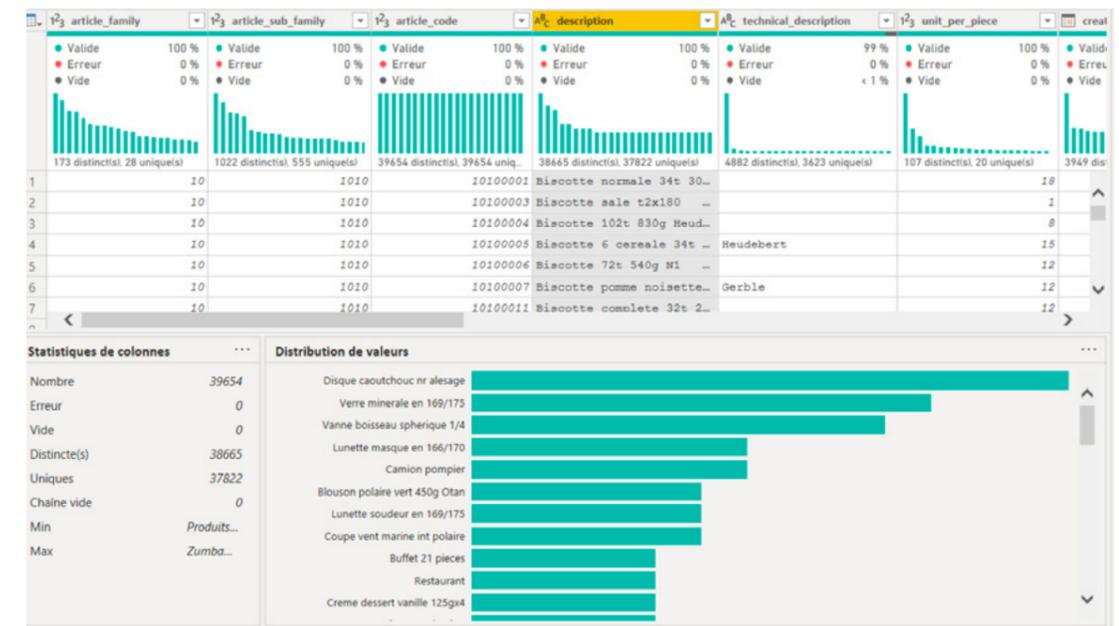
Ces points de repère généraux étant posés, nous proposons ici, plus qu'un « benchmark » des solutions possibles, qui contreviendraient à l'impératif d'adaptation au contexte et la maturité ambiante à un instant t, un ensemble de questions à se poser avant d'envisager l'acquisition d'un ou plusieurs outils dédiés à la gestion de la qualité des données :

- Quelles fonctionnalités de gestion de qualité des données sont déjà couvertes par les outils et le SI existant ?
- Quel apport de valeur voit-on à combler les manques en termes de fonctionnalité, ou à remplacer tout ou partie des fonctionnalités par une nouvelle solution (outil) ?
- Est-on certain que la maturité et le modèle opérationnel actuel permettent de faire une transition vers la nouvelle solution (outil) ?
- En corollaire, a-t-on bien anticipé la nécessaire conduite du changement associée à l'adoption de la solution (outil) ?
- Existe-t-il des facteurs internes qui vont faciliter l'intégration technique de la nouvelle solution, ou au contraire la rendre plus difficile ?
- Existe-t-il des facteurs internes qui vont permettre de réduire le Coût Total de Possession (Total Cost of Ownership) - par exemple, présence d'un ETL chez le même éditeur ?
- Comment voit-on l'évolution projetée en termes d'utilisation et d'adoption de la solution (périmètre métier, périmètre technique, nombre d'utilisateurs, etc.) ?
- Opère-t-on dans un cadre réglementaire qui inciterait à l'utilisation d'une telle solution (outil) ?
- De quel budget dispose-t-on ? Quelle part ce budget représente-t-il par rapport aux autres engagements budgétaires nécessaires pour le Data Management / Data Governance ?

Exemple de profilage réalisé avec l'outil Informatica IDQ (suite DQM complète)



Exemple de profilage réalisé avec un outil de dataviz (PowerBI)



OUVERTURE SUR DE NOUVELLES APPROCHES

7

OUVERTURE SUR DE NOUVELLES APPROCHES POUR ANALYSER LA QUALITÉ / IDENTIFIER DES ANOMALIES

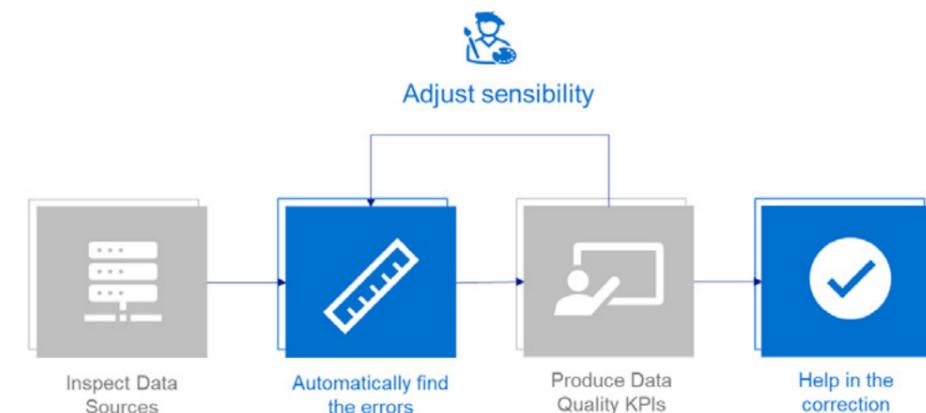
Comme nous venons de le voir, l'outillage est un atout structurant pour répondre aux besoins de documentation, d'évaluation de la qualité et de pérennité du dispositif dans un environnement SI qui évolue perpétuellement. Malgré les fonctionnalités de plus en plus sophistiquées, ces solutions proposent des visions que nous pouvons qualifier de statiques et qui reposent généralement sur des périmètres ou critères (pré)définis lors du projet « Qualité des données ».

Les dispositifs « Quality by Design » répondent de manière plutôt satisfaisante aux enjeux d'actualisation de la documentation et des règles de lignages toutefois des limites existent concernant les nouveaux périmètres. Les algorithmes d'intelligence artificielle ne pourraient-ils pas apporter une vision plus dynamique, plus contextualisée de la qualité des données ?

Il est souvent constaté, lors des projets qualité de données, des difficultés dans la définition des règles ou des critères qualifiants par les équipes métiers/utilisateurs. En effet, partir d'une page blanche ou d'exemples très simples, laisse souvent les contributeurs songeurs ou dubitatifs devant cet exercice. Dans ce cas, les algorithmes intelligents peuvent mettre en évidence les valeurs atypiques/anormales en fonction de corrélations observées sur 1 à n champs et pour l'ensemble de l'historique des données. Les combinaisons majoritaires (ou très fréquentes) sont ainsi considérées comme des combinaisons de référence et les atypismes sont isolés pour analyse. Les combinaisons considérées comme finalement conformes seront validées manuellement et reconnues par l'algorithme ; les autres seront mémorisées comme règles d'isolement. Cette technique peut être utilisée :

- en première analyse pour accompagner les métiers à définir/qualifier les règles et critères de qualité
- en mode run pour enrichir et actualiser les nouvelles règles développées. Les nouveaux cas d'usage seraient isolés dès leur apparition, nécessitant l'analyse des différents protagonistes, au plus tôt du cycle de vie des (nouvelles) données.

Champs d'application privilégiés de l'IA pour la gestion de la qualité des données



La capacité de traitements pourra aussi être utilisée pour effectuer des rapprochements de masse entre les données « produites » et des données de référence internes ou externes, ou la détection de biais potentiel d'un jeu de donnée entre deux périodes.

En synthèse, l'IA permet d'enrichir sa pratique Qualité des Données sur plusieurs dimensions.

Dimension	Description	Que peut apporter l'IA ?	Exemple de techniques pertinentes
Complétude	Donnée renseignée, pas de donnée manquante	Automatiser le process de traitement (identification puis remplissage) des valeurs manquantes	• Recommandation « Reinforcement Learning »
Intégrité	Lors de son cycle de vie la donnée n'est pas corrompue ou modifiée	N/R	-
Fraicheur	Données à jour ou disponible au moment opportun	N/R	-
Unicité	Donnée non redondée dans une source	Identifier des profils similaires (« fuzzy matching ») pour dédupliquer une base de données	• Clustering • Classification
Exactitude	La donnée correspond à la réalité qu'elle décrit	Comparer et qualifier la qualité de l'information en la croisant avec des sources externes	• Détection d'Anomalie • "Deep Learning" (vision, NLP – Natural Language Processing)
Vraisemblance	Donnée plausible, dans un éventail de valeurs crédibles ou ayant du sens dans son contexte d'utilisation	Confirmer la vraisemblance en construisant un profil de la donnée	• Régression • Classification • Détection d'Anomalie
Cohérence	La valeur de la donnée ne diverge pas avec la valeur d'autres données avec lesquelles elle est corrélée	Détecter des corrélations subtiles	• Régression • Détection d'Anomalie
Validité	Donnée appartenant à une liste de valeurs prédéfinies	N/R	-

L'IA n'est cependant pas une solution miracle, et il faut garder en tête quelques contraintes de mise en œuvre :

- l'IA peut faciliter les contrôles, voire suggérer des pistes de remédiation (après entraînement ad hoc) , mais elle ne va pas aider dans l'identification des causes racine.
- Le paramétrage d'un modèle performant nécessite de disposer du volume suffisant de données labellisés pour recourir à des approches supervisées.
- Le déploiement dépend de la confiance des collaborateurs dans le modèle : les modèles sans explications (« black box ») sont à éviter, de même que la génération de trop de faux positifs (cas fréquent en anomaly detection).

En résumé, l'utilisation des algorithmes peut être un facilitant pour la recherche et la quantification des problèmes ou anomalies, ainsi que dans la correction leur correction. L'IA est alors particulièrement efficace lorsque l'aspect multi-dimensionnel des variables à traiter est bien marqué. Cependant, elle ne doit en aucun cas se substituer à l'analyse ou l'implication des équipes métiers au cours du dispositif qualité des données. Mais les progrès en la matière étant particulièrement rapides, il conviendra de nous redonner rendez-vous dans quelques années sur l'état de l'art de l'IA au service de la qualité des données.

n°2 POINTS DE REPÈRE CLÉS POUR MESURER LA QUALITÉ DES DONNÉES



DAMA
FRANCE



ISACA

AFAI | PARIS-FRANCE CHAPTER



Groupe de Travail

QUALITÉ DES DONNÉES

cahier
pratique

n°
2

**POINTS DE REPÈRE CLÉS
POUR MESURER LA QUALITÉ
DES DONNÉES**



Seconde édition Publiée à Paris le 30 Novembre 2022
par ISACA-AFAI & DAMA France

Direction éditoriale : Virginie Cuvilliez, Marcel Lee, Vincent Manière et Mathieu Minault
Conception graphique : Agence Imedia Conseil
Crédits photographiques Shutterstock et AdobeStock

© 2022 - Tous droits réservés

Groupe de Travail

QUALITÉ DES DONNÉES



ISACA

AFAI | PARIS-FRANCE CHAPTER

DAMA
FRANCE