



Management des données

Gestion des données – La qualité des données

elie_chemaly@yahoo.fr



Introduction

En licence, vous avez vu que les systèmes d'informations sont un système complexe qui mélange les acteurs (humains / non humains), les territoires (physiques, virtuels) et les projets (qui aligne l'informatisation des territoires sur les besoins des acteurs).

Aujourd'hui vous êtes au premier semestre de Master 1, et vous avez commencé par une certification TOSA. Cette certification permet de comprendre et maîtriser un outil comme Excel. Cet outil est utilisé massivement dans les entreprises pour automatiser des tâches de traitement de données (reporting, statistiques, graphiques, vérification de données, contrôle, etc.).

C'est donc un outil qui permet d'informatiser le territoire de beaucoup d'acteurs.

Le problème qui se pose est le suivant : les données qui sont manipulées dans ces nombreux (voir très nombreux) fichiers Excel ne sont pas contrôlées : D'où viennent-elles ? Qui les a produites ? Est-ce qu'elles sont fiables ? ...

Autre exemple : Dans l'audit, pour contrôler les comptes ou les activités, vous allez utiliser des données et certainement Excel. C'est bien, il faut donc connaître Excel, mais il vous manque la partie de compréhension sur la gestion des données.

D'où ce cours, qui vous présente la façon dont une entreprise, et en général la direction des systèmes d'information, met en place une démarche globale de gestion des données.

Après ce cours, au semestre 2, vous aurez un cours dédié à l'Audit du système d'information avec Fabien Briard. Ce cours est donc en amont pour vous présenter la gestion des données d'entreprises

La qualité données

Quiz

Qualité pas qualité

La qualité données

Qu'est ce que la qualité des données?

La qualité des données peut être considérée comme la somme de toutes les caractéristiques des données qui déterminent leur utilité dans l'exécution de processus métiers, décisionnels ou non (BLM, novembre 2002).

Sans une qualité des données optimale, les processus métiers s'en trouvent donc altérés et les processus décisionnels se fondent sur de mauvais indicateurs.



La qualité des données représente le niveau de conformité des caractéristiques intrinsèques d'une ou plusieurs données par rapport aux exigences définies par le consommateur et acceptées par le producteur dans un contexte d'utilisation. Nous pouvons attribuer l'adjectif "Pauvre", "Bien" ou "Excellent" à un niveau de qualité d'une donnée.

La qualité données

A quoi bon va me servir la qualité des données



- Quand j'effectue un audit des comptes d'une société, si les fichiers de données qui sont mis à ma disposition sont erronées (pas à la bonne période, illisibles, pas de confiance), les efforts que je pourrais déployer pour fournir un travail de qualité peuvent s'avérer vains → Je perds mon temps, le client est insatisfait et l'utilisation du résultat de mon travail peut induire en erreur
- Les écritures comptables des charges d'une société n'ayant pas de document à valeur probante ne sont pas conformes et doivent être reconsidérées ou corrigées.
- Quand je demande de me fournir des documents pour effectuer un contrôle, et la clé usb fournie est illisible

La qualité données

Quand considère-t-on une donnée de qualité?



Une donnée est de qualité quand elle correspond aux objectifs pour laquelle elle a été produite dans un contexte d'utilisation et dans un lapse de temps.

Exemple : les mesures de la longueur et de la largeur d'un mur destinées à un peintre en vue d'estimer les travaux de peintures, ne peuvent pas servir pour un menuisier pour la fabrication d'une bibliothèque qui va longer le même mur. La précision exigée par le menuisier pour fabriquer son meuble est bien plus élevée que celle du peintre.

La qualité données

Quand considère-t-on une donnée de qualité?



Une donnée est de qualité quand elle présente une conformité syntaxique par rapport à son format déclaré

Une donnée de format date AAAAMMJJ ne peut pas servir pour un montant financier

La qualité données

Quand considère-t-on une donnée de qualité?



Une donnée est de qualité quand elle fraîche.

Lorsque vous effectuez des achats en ligne, les prix affichés sont les plus récents définis par le vendeur pour ses produits.

La qualité données

Quand considère-t-on une donnée de qualité?



Une donnée est de qualité quand elle est intègre par rapport à la donnée à laquelle elle se réfère.

Un vendeur en ligne souhaite effectuer un remboursement pour un client ayant retourné ses articles achetés, il faut s'assurer de l'utilisation des coordonnées bancaires du client en question et non pas ceux d'un autre client.

La qualité données

Quand considère-t-on une donnée de qualité?



Une donnée est de qualité quand le réceptacle qui la stocke est de qualité.

Vous disposer d'une clé USB contenant l'intégralité de vos cours. Elle vous sert au quotidien pour consulter les documents qui y sont stockés. Si elle devient illisible par erreur de manipulation, elle a beau héberger tous vos cours, elle est inutilisable.

La qualité données

Quand considère-t-on une donnée de qualité?



Une donnée est de qualité quand les règles des traitements qui la calculent sont de qualité.

Le montant total d'une facture de marchandise doit prendre en compte l'intégralité des articles achetés, les quantités respectives, les remises, les taxes... Il est important de s'assurer de la justesse de l'intégralité des opérations de calcul ayant servis au calcul du montant total.

La qualité données

Quand considère-t-on une donnée de qualité?



Une donnée est de qualité quand elle reflète les faits qu'elle représente dans un espace temps défini.

Les prix définis pour une promotion sur une sélection de parfums est limité sur une période. Promotion du 22/11/2022 → 28/11/2022. Passé ce délai, la remise n'est plus applicable.

La qualité données

Quand considère-t-on une donnée de qualité?



Une donnée est de qualité quand son canal de communication préserve son intégrité depuis son acquisition jusqu'à sa diffusion.

Vous êtes en communication téléphonique et vous traversez une zone mal couverte par le réseau GSM. Vous entendez toujours la voix de votre interlocuteur hachée cependant vous ne comprenez plus ses messages.

La qualité données

Quand considère-t-on une donnée de qualité?



Les données produites sont ainsi en qualité quand elles correspondent à l'usage dans les conditions définies par les métiers et les consommateurs et acceptées par le producteur.

La qualité données

Qu'est ce qu'une exigence de qualité des données, comment elle se traduit opérationnellement?

Une **exigence** de qualité est la définition détaillée de la qualité des données demandée par les métiers responsables des données concernées.



Elle est exprimée sur la base des dimensions de mesure : intégrité, complétude, cohérence, unicité, validité, temporalité.

Les exigences de qualités sont définies dans les spécifications, et traduites opérationnellement par des règles de contrôle de qualité dans les développements

La qualité données

Quels sont les conséquences de la non qualité des données?

Les données sont considérées comme actifs immatériels de l'entreprise

Préserver la qualité de ses données permet de protéger et augmenter la valeur du patrimoine données d'une organisation.



La non-qualité des données expose l'entreprise d'aujourd'hui à des risques : financiers, perte de productivité, réglementaires, opérationnels, déficit d'image.

La qualité données

Exemple d'impact suite à un problème de qualité de données

Cas : Investissements inutiles entamés sur la base d'indicateurs calculés avec des données de mauvaise qualité.



Le directeur d'une grande banque de la place, fort de plus d'un million de comptes clients dans son portefeuille, a entrepris une démarche d'élargissement de son réseau d'agences par le recrutement du personnel nécessaire, de l'achat, de la location de locaux, et du renforcement de son système d'information.

Un contrôle de qualité effectué dans le cadre d'une migration de données des clients vers le nouveau système a mis en exergue le vrai nombre de clients : **600 000 clients uniques** au lieu **de plus d'un million estimé**.

En effet, le million de comptes clients appartiennent réellement à **600 000 clients uniques**.



→ Une banque a effectué des investissements d'élargissement de son réseau d'agences en se basant sur des indicateurs **calculés** à partir de données de mauvaises qualités. Les investissements s'avèrent inutiles après la découverte du vrai nombre de clients.

La qualité données

Quiz

La qualité données

Comment mesurer la qualité des données

Il est important de s'appuyer sur des unités de mesure objectives afin d'interpréter un niveau de qualité d'une donnée. Les unités de mesure de la qualité des données sont dites dimensions de mesure de qualité de données.

1

COMPLÉTITUDE



2

UNICITÉ



3

TEMPORALITÉ



4

VALIDITÉ



5

COHÉRENCE



6

INTÉGRITÉ



La qualité données

Les dimensions de mesure de la qualité des données

1



COMPLÉTUDE

Elle représente le taux de remplissage des données dans un gisement de données existant.

Elle est mesurée par un contrôle de valorisation des champs à vérifier, ou par le contrôle de complétude d'une structure de données par rapport aux attributs définis dans la spécification. Exemple : vérification de l'existence de l'intégralité des champs dans une table par rapport à la spécification.

La qualité données

Les dimensions de mesure de la qualité des données

2



L'UNICITÉ

Elle consiste à vérifier l'unicité des instances d'une même donnée réputée unique dans un gisement de données.

Elle est vérifiée par un contrôle de vérification de doublons dans un gisement de données sur une donnée définie unique. Exemple : la vérification de l'unicité d'une personne physique dans un référentiel de consommateurs.

La qualité données

Les dimensions de mesure de la qualité des données

3



LA TEMPORALITÉ

La temporalité de la donnée correspond à la vérification de l'adéquation de la donnée par rapport aux exigences des métiers en termes de fraîcheur.

Elle est mesurée par un contrôle de vérification de date de création ou date de dernière modification ou date de mise à disposition des données par rapport à une date de référence, ou une durée de fraîcheur de référence. Exemple : vérification de la date de dernière relève manuelle.

La qualité données

Les dimensions de mesure de la qualité des données

4



LA VALIDITÉ

La validité d'une donnée correspond à l'adéquation des données aux faits réels.
La validité est vérifiée essentiellement en mode manuel par un contrôle humain.
Exemple: vérification de l'identité d'une personne physique, vérification de l'adresse réelle d'une personne physique..

La qualité données

Les dimensions de mesure de la qualité des données

5



LA COHÉRENCE

Elle consiste à mesurer la conformité d'une donnée existante par rapport à une autre donnée dans le même gisement ou dans un gisement externe.

Elle est mesurée par une comparaison entre deux données. Exemple: la vérification de l'adéquation entre la date de naissance et l'âge, ou la vérification de la validité de l'adresse d'une personne physique par rapport à un référentiel d'adresses externes.

La qualité données

Les dimensions de mesure de la qualité des données

6



L'INTÉGRITÉ

C'est une dimension technique qui consiste à vérifier que la donnée est préservée lors des opérations de transports et de mise à disposition, et qu'elle n'a pas été ni altérée, ni volée. L'intégrité de la donnée peut être aussi liée à la modélisation que l'on appelle aussi intégrité référentielle.

L'intégrité est mesurée par la comparaison de la taille en amont et en aval de l'opération de transfert, par les contrôles d'intrusion réseau.