le cnam

CONSERVATOIRE NATIONAL DES ARTS ET METIERS
PARIS

MEMOIRE

présenté en vue d'obtenir

UE « Information et Communication pour Ingénieur »

SPECIALITE : Informatique
OPTION : Architecture et Ingénierie des Systèmes et logiciels

par

Saulnier Jean-Pierre

Le Master Data Management :
rôle et enjeux dans la performance des entreprises

Soutenu le 26 juin 2018

JURY

PRESIDENT : Monsieur Philippe RIGAUX, Professeur au CNAM

MEMBRES : Madame Françoise SAILHAN, Maître de conférence au CNAM
Monsieur Eric SOUTIL, Maître de conférence au CNAM

Remerciements

Je tiens à remercier Madame ARRUABARRENA et Madame CHARTRON pour la qualité des cours prodigués et les conseils donnés.

Je remercie toutes les personnes qui ont accepté de me relire et de me faire part de leur avis.

Merci également à l'équipe de France Télévisions pour leurs critiques constructives lors de ma présentation.

Et bien sûr merci à Monsieur Rigaux d'avoir validé mon plan.

Glossaire

Bases de données relationnelles : base de données, ayant une ou plusieurs relations entre elles, où l'information est organisée dans des tableaux à deux dimensions appelés des tables. MySQL en est un exemple courant.

Clef primaire : donnée qui permet dans une base de données relationnelle d'identifier de manière unique un enregistrement dans une table.

Clef étrangère : contrainte dans une base de données relationnelle qui garantit l'intégrité référentielle entre deux tables. Elle a généralement comme valeur un nombre faisant référence à la clef primaire de l'objet lié (une commande peut avoir comme clef étrangère fac_id dont la valeur correspondra à la clef primaire d'une facture de la table facture, par exemple).

Data mining : famille d'outils facilitant l'exploration et l'analyse des données contenues au sein d'une base décisionnelle de type entrepôt de données.

La compétitivité-prix : désigne la capacité d'un produit à attirer des clients au détriment des produits concurrents du fait de son prix. Sa mesure permet de situer la place de l'entreprise sur le marché par rapport à ses concurrents.

La compétitivité hors-prix : désigne la capacité d'un produit à attirer des clients au détriment des produits concurrents du fait des éléments indépendants du prix. Elle est obtenue grâce à des éléments comme la qualité des produits, l'innovation, le service, le design…

Client/serveur : mode de communication permettant à des utilisateurs finaux d'être connectés à travers un réseau à des serveurs via leur ordinateur.

Enterprise Information Management (EIM) : regroupe l'ensemble des processus, technologies et organisations nécessaires pour transformer et réconcilier sémantiquement les données afin de garantir l'intéropérabilité des systèmes.

Entrepôt de données : « data warehouse » en anglais. Base de données spécifique issue de plusieurs autres sources et réservée à l'aide à la prise de décision (Business Intelligence).

Extract Transform Load (ETL) : technique permettant d'effectuer des synchronisations massives de données d'une base de données vers une autre.

Index : un index est utilisé et entretenu par le système de gestion de base de données pour lui permettre de retrouver rapidement les données. Par exemple, la clef primaire est un index naturel.

Matrices de Porter, Mac Kinsey : matrices stratégiques permettant une analyse globale et simultanée de l'ensemble des activités d'une entreprise par rapport au marché.

Organisation Internationale de Normalisation : Organisme ayant pour but de produire des normes internationales, dans les domaines industriels et commerciaux, appelées normes ISO.

Processus : désigne la succession des tâches qui concourent à l'élaboration d'un produit – qu'il s'agisse d'un bien, d'un service ou d'un assemblage de biens et de services.

Profilage des données : processus qui consiste à récolter les données dans les différentes sources de données existantes (bases de données, fichiers, etc.) et à collecter des statistiques et des informations sur ces données.

Progiciel de Gestion Intégré (ERP) : ensemble de logiciels intégrant les principales fonctions nécessaires à la gestion des flux et des procédures de l'entreprise (comptabilité et finances, logistique, paie et ressources humaines, etc.)

Référentiel : base dans laquelle l'entreprise documente une partie de ses règles de fonctionnement, technique ou fonctionnelles. Ces informations peuvent décrire comment est structuré le Système d'Information, selon quelles règles les données de l'entreprise sont transférées et transformées d'une base à une autre, comment se décomposent les différents processus…

Reporting : application permettant de sélectionner des données relatives à une période donnée pour évaluer le niveau de production d'un produit pour un secteur de clientèle, de trier, regrouper ou répartir ces données selon des critères choisis, de réaliser divers calculs (totaux, moyennes, écarts, comparatif d'une période à l'autre…), de présenter les résultats d'une manière synthétique ou détaillée selon leurs besoins ou les attentes des dirigeants de l'entreprise.

Service Oriented Architecture (SOA) : modèle d'interaction applicative mettant en œuvre des services (en général des services web échangeant en JSON ou XML).

Système opérationnel : système dédié aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes des commandes, de la logistique, de facturation, des comptes clients, de fidélisation, etc.

Système d'enregistrement distribué : référentiel de coopération situé en milieu de chaîne assurant l'unicité de l'information par le recueil d'informations d'identification de la donnée enregistrée par différents systèmes source.

Trigger : événement qui provoque l'exécution d'un algorithme (la mise à jour d'une ligne en base de données relationnelles, par exemple)

Urbanisation : ensemble des moyens permettant de faire évoluer le Système d'information et le système informatique aux mêmes rythmes que la stratégie et l'organisation. Elle consiste à décrire la structuration du système cible et la façon de l'atteindre.

Services web : « web services » en anglais. Mécanisme de communication entre applications distantes à travers le réseau internet, indépendant de tout langage de programmation et de toute plate-forme d'exécution. Il utilise le protocole HTTP comme moyen de transport ainsi qu'une syntaxe basée sur XML ou JSON.

Abréviations

BCG : **B**oston **C**onsulting **G**roup. Outil qui permet d'évaluer l'équilibre du portefeuille d'une entreprise et qui permet de positionner les activités selon deux dimensions : le taux de croissance du marché et la part de marché relative détenue.

EVA : création de valeur intrinsèque (« **E**conomic **V**alue **A**dded » en anglais). Ce ratio permet de mesurer la création de valeur pour l'actionnaire. C'est la différence entre le résultat opérationnel et les capitaux investis.

JSON : **J**avaScript **O**bject **N**otation. Format léger d'échanges de données facile à lire et à écrire pour des humains, se basant sur la structure clef/valeur.

PHP : **P**HP **H**ypertext **P**reprocessor. Il s'agit d'un un langage de scripts généraliste et Open Source, spécialement conçu pour le développement d'applications web. Il peut être intégré facilement au HTML. Actuellement dans sa version 7.

ROE : retour sur les attentes (« **R**eturn **O**n **E**xpectations » en anglais). Ce ratio mesure la rentabilité financière des capitaux apportés par les propriétaires de l'entreprise. C'est le rapport entre le résultat net et les capitaux propres.

ROI : retour sur investissement (« **R**eturn **O**n **I**nvestment » en anglais). Ce ratio mesure la rentabilité économique du capital utilisé par l'entreprise. C'est le rapport entre le résultat d'exploitation et les capitaux investis.

SWOT : de l'anglais « **S**trengths » (forces), « **W**eaknesses » (faiblesses), « **O**pportunities » (opportunités), « **T**hreats » (menaces). Outil de stratégie d'entreprise permettant de déterminer les options offertes dans un domaine d'activité stratégique.

XML : e**X**tensible **M**arkup **L**anguage. Il s'agit d'un langage informatique de balisage générique. (<title></title> par exemple)

Table des matières

Remerciements ..... **- 2 -**
Glossaire ..... **- 3 -**
Abréviations ..... **- 5 -**
Introduction ..... **- 7 -**
1) La performance des entreprises ..... **- 8 -**
2) Le Système d'Information au service de la performance ..... **- 9 -**
2.1) Définitions ..... *- 9 -*
2.1.1) La donnée ..... - 9 -
2.1.2) Le système d'Information ..... - 10 -
2.2) Objectifs ..... *- 11 -*
2.2.1) Les objectifs opérationnels ..... - 11 -
2.2.2) Les objectifs d'analyse et d'aide à la décision ..... - 11 -
2.3) Les problématiques du Système d'Information ..... *- 12 -*
2.3.1) La spécialisation des données échangées ..... - 12 -
2.3.2) Les contraintes internes et externes du Système d'Information ..... - 13 -
2.4) L'enjeu de la performance ..... *- 14 -*
3) La Gestion de la Qualité des Données ..... **- 15 -**
3.1) Définition ..... *- 15 -*
3.2) Les critères de qualité ..... *- 16 -*
3.2.1) Les critères intrinsèques ..... - 16 -
3.2.2) Les critères de service ..... - 16 -
3.2.3) Les critères de sécurité ..... - 17 -
3.3) L'impact négatif sur la performance ..... *- 17 -*
3.4) L'impact positif sur la performance ..... *- 18 -*
3.5) Un besoin d'urbanisation du Système d'Information ..... *- 19 -*
4) Les données de référence ..... **- 20 -**
4.1) Définition et typologie ..... *- 20 -*
4.2) Dictionnaires de données ..... *- 21 -*
4.2.1) Métadonnées métier ..... - 21 -
4.2.2) Métadonnées techniques ..... - 22 -
4.2.3) Métadonnées administratives ..... - 23 -
4.3) Nécessité d'une gestion des données de référence ..... *- 23 -*
5) Gestion des Données de Référence ..... **- 24 -**
5.1) Définition et évolution du périmètre des données traitées ..... *- 24 -*
5.2) Fonctions principales assurées par l'architecture Master Data Management ..... *- 24 -*
5.2.1) L'intégration des données de référence ..... - 24 -
5.2.2) La gouvernance des données de référence ..... - 25 -
5.2.3) La qualité des données de référence ..... - 26 -
Conclusion ..... **- 28 -**
Bibliographie ..... **- 30 -**
Liste des figures ..... **- 32 -**
Liste des tableaux ..... **- 32 -**

Introduction

Pour assurer sa pérennité et ses emplois, une entreprise doit veiller à sa performance dans tous les domaines : financier, technique, innovation, commercial... Un certain nombre de mesures permettent de l'évaluer en se basant sur les données du Système d'Information.

Celui-ci, soumis à des contraintes légales, répond à des besoins opérationnels, d'analyse et d'aide à la décision. Pour cela, il doit s'appuyer sur des données de qualité.

Or, ces données, échangées entre les différents services du Système d'Information, se spécialisant de plus en plus, deviennent hétéroclites, éparses, incohérentes, obsolètes. Elles peuvent alors être source d'erreurs, perturber son bon fonctionnement et se révéler fatales pour l'avenir de l'entreprise.

C'est la raison pour laquelle certaines données, dites de référence (ou master data), doivent être gérées. Nous étudierons comment la Gestion des Données de Référence (ou Master Data Management) joue un rôle dans le contrôle de leur qualité afin d'améliorer la performance de l'entreprise tant aux niveaux opérationnel que décisionnel.

1) La performance des entreprises

La performance se mesure avec des critères ou indicateurs qualitatifs ou quantitatifs de résultat permettant d'évaluer son efficacité et son efficience. Ces indicateurs sont un outil d'aide à la décision et à la mesure de la bonne santé de l'entreprise.

Touchant à tous les domaines d'activité de l'entreprise, ils permettent de connaître l'efficacité de la production, renseignent sur la qualité de la relation commerciale et du service client, permettent d'évaluer le temps passé à corriger des erreurs, mesurent la perception de l'entreprise et fournissent des informations sur la qualité des services.

Pour mesurer l'efficacité, on utilise un critère qui exprime un rapport entre le résultat obtenu et l'objectif visé (le nombre de ventes d'un produit par rapport aux prévisions estimées sur une période donnée, par exemple). L'efficience est quant à elle mesurée par un critère qui exprime un rapport entre le résultat financier obtenus et les ressources financières mis en œuvre pour produire ces biens et services. [19]

Pour évaluer la performance d'une entreprise, il est nécessaire d'effectuer des mesures à tous les niveaux : financier, économique, organisationnel, social et sociétal.

La performance financière : on mesure la performance financière à l'aide des indicateurs de retour sur investissement (ROI), retour sur les attentes (ROE) et création de valeur intrinsèque (EVA). L'Excédent Brut d'Exploitation est également couramment utilisé.

La performance économique : elle concerne la compétitivité-prix et la compétitivité hors-prix. Les indicateurs permettant de la mesurer sont les taux de transformation d'un devis en facture, le chiffre d'affaire dégagé par les nouveaux produits ou services, les indicateurs de notoriété sur les réseaux sociaux, les taux de fidélité, le nombre de clients perdus sur une période, etc.

La performance organisationnelle : il s'agit de mesurer la performance de l'entreprise au niveau de la qualité de la production, des délais, des procédures métiers, etc. Des indicateurs peuvent être le temps passé aux procédures administratives ou le nombre de contacts émis par un collaborateur chaque jour, par exemple.

La performance sociale : le bilan social récapitule les principales données chiffrées permettant d'apprécier la performance sociale et les rapports sociaux au sein d'une entreprise. Parmi les nombreux indicateurs sociaux, on peut citer : le montant des rémunérations, le nombre d'accidents de travail, les maladies professionnelles, le taux de formation des employés, le taux d'absence et de démission des employés, etc.

La performance sociétale : indique l'engagement de l'entreprise dans les domaines environnementaux, humanitaires, culturels. Les outils de la responsabilité sociétale de l'entreprise (RSE) peuvent être utilisés pour apprécier le niveau de performance de l'entreprise.

Ces indicateurs de pilotage (statistiques logistiques, de ventes, de finance et d'achats, etc.), définis sur des tableaux de bord et de reporting, sont créés par le système décisionnel qui se base sur des données créées par les processus transactionnels (achats, ventes, production, logistique, etc.) du Système d'Information.

2) Le Système d'Information au service de la performance

2.1) Définitions

2.1.1) La donnée

Une donnée est une description élémentaire de nature numérique d'une réalité (chose, événement, mesure, transaction, etc.), représentée sous forme codée, en vue d'être collectée, traitée, conservée, échangée [1]. Elle peut être valorisée et a un cycle de vie [2].

La donnée peut être **structurée**, **semi-structurée** ou **non-structurée**. [1] [3]

Une donnée structurée est une donnée dont on a établi fonctionnellement le sens de manière détaillée, son cycle de vie et donc ses règles de création, les valeurs possibles dans le cas de listes, ainsi que le moyen technique de représentation. Les données structurées sont stockées ou échangées dans un format organisé, une syntaxe, et dont la sémantique a préalablement été définie (par exemple une table dans une base de données ou un document XML).

Les données non-structurées sont également une description d'une réalité, mais dont la codification, dont le sens, n'est pas exploitable directement par la machine : un fichier audio, vidéo, un texte contenu dans un document ou un mail par exemple.

Des données semi-structurées combinent les deux types : une partie structurée, et une partie non structurée. Certaines informations peuvent leur être associées, telles que des balises de métadonnées, qui permettent l'adressage des éléments qu'elles renferment (un document Word à qui l'on a rajouté des métadonnées sous la forme de mots-clés représentant le contenu du document et permettant de le retrouver plus facilement lorsqu'une recherche est effectuée sur ces termes, par exemple). [4]

2.1.2) Le système d'Information

Avant de définir ce qu'est un Système d'Information (SI), il est important ici de rappeler la relation entre information et donnée, l'une découlant de l'autre. En effet, selon le document de référence de l'urbanisation du Système d'Information de l'Etat, « une information est un ensemble de données agrégées en vue d'une utilisation par l'homme » [1] et est compréhensible par un humain. Une donnée est, quant à elle, brute, et pas forcément compréhensible par un humain.

Un système d'information peut se résumer avec le schéma suivant [5] :

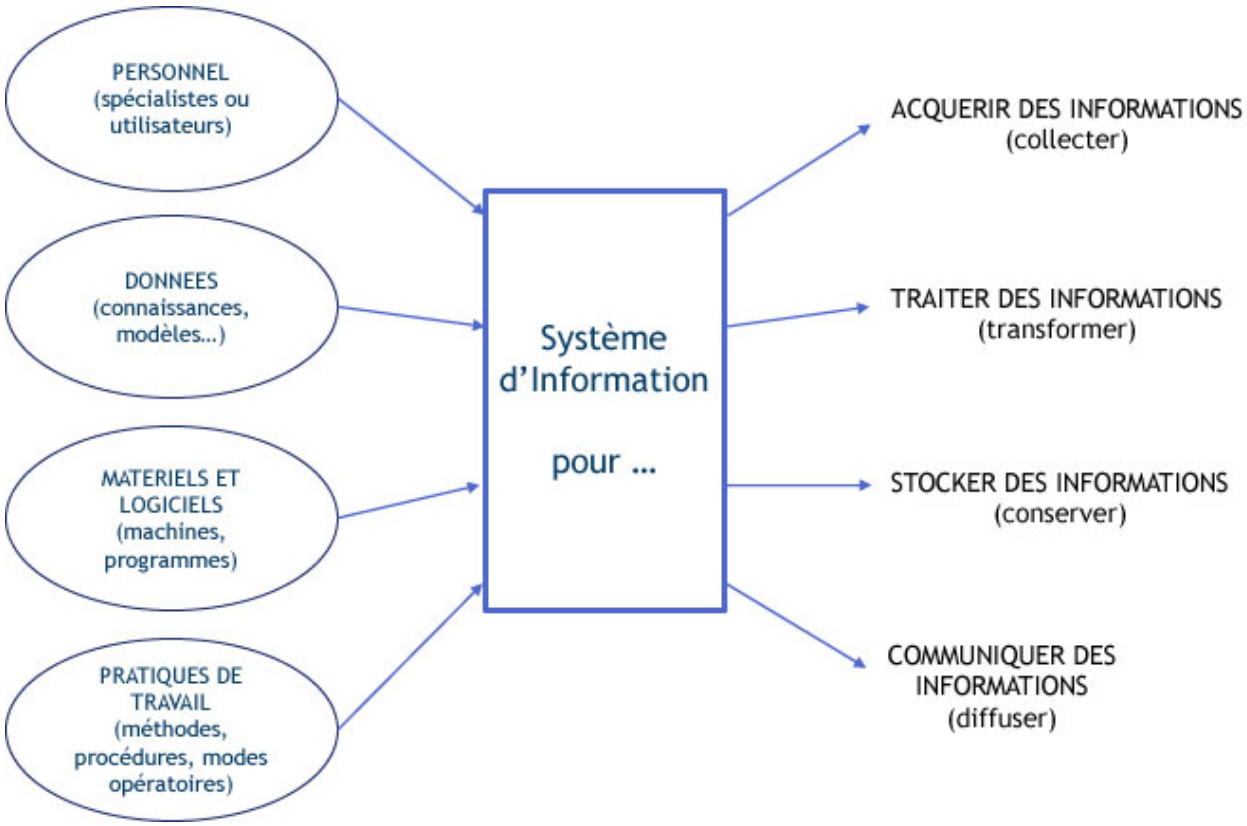


Fig 1 - Le Système d'Information [7]

2.2) Objectifs

Les objectifs d'un Système d'Information sont [5] :

- opérationnels
- stratégiques

2.2.1) Les objectifs opérationnels

La plupart des acteurs de l'entreprise ont besoin d'informations et manipulent des données pour exercer leur métier : assistants commerciaux, acheteurs, responsables des stocks, comptables, responsables des ventes, clients ou prospects, etc.
Ces besoins nécessitent de saisir, consulter et diffuser des données ainsi que d'éditer des documents (formulaires, listes).

Un grand nombre de données sont partagées par plusieurs métiers. [3][6]
Les clients, par exemple, sont segmentés par industrie, rattachés à des équipes commerciales, liés à des comptes financiers, gérés par des organisations, rattachés à des géographies, paient en devises.
De même, les produits sont segmentés par industrie, vendus par des équipes commerciales, liés à des comptes financiers, produits et vendus par des organisations, vendus dans des géographies et tarifés en devises.

Les métiers de l'entreprise, pour pouvoir être exercés correctement, ont besoin que l'information adéquate soit disponible au bon moment avec le niveau de granularité requis. [6][7]

Il est donc important que ces données ne soient pas redondantes. C'est l'une des raisons qui poussent les entreprises à adopter un progiciel de gestion intégré (ERP), qui ne garantira cependant pas des données de qualité.

2.2.2) Les objectifs d'analyse et d'aide à la décision

Afin d'analyser l'activité de l'entreprise, les managers suivent des tableaux de bord de pilotage et de reporting qui sont basés sur des indicateurs précis et pertinents (le chiffre d'affaires, le nombre de ventes ou le montant de taxes pour les ventes de produits par répartition géographique/magasins par exemple), leur permettant de prendre rapidement les bonnes décisions. [9]

Ces indicateurs se basent sur des données opérationnelles stockées dans des bases de données, des fichiers plats, des services web et in fine dans des entrepôts de données (« data warehouse » en anglais). Ils sont construits selon un ou plusieurs axes d'analyse. Par exemple, l'analyse des achats de matières premières peut se faire par famille, par article, par fournisseur, par mois. [6]

Ces indicateurs de pilotage sont principalement produits à partir de données issues du système opérationnel. Par exemple, le chiffre d'affaires d'une entité pour un mois donné sera obtenu à partir des commandes enregistrées par cette entité sur la période considérée.

Les managers peuvent ainsi mesurer les effets et performance de leurs décisions, avoir à leur disposition les informations nécessaires pour le développement d'une stratégie et d'un plan d'action (par les pratiques de data mining par exemple) et identifier les faiblesses et les forces de l'organisation (par les matrices d'analyse stratégiques, par exemple : BCG, Porter, SWOT, etc.)

2.3) Les problématiques du Système d'Information

2.3.1) La spécialisation des données échangées

Si, à l'origine, les données étaient cloisonnées dans le SI de chaque service pour des raisons de commodité et de réactivité, aujourd'hui, les différents services ont besoin de communiquer entre eux des données qu'ils partagent. Cependant, le format des données qu'ils échangent pouvant être différent, des interfaces sont généralement mises en place par les entreprises pour pouvoir communiquer. Le souci lié à ce genre de pratique est qu'à chaque nouvel échange, une nouvelle interface doit être créée, ce qui entraîne ce que Gartner Group appelle le « syndrome spaghetti ».

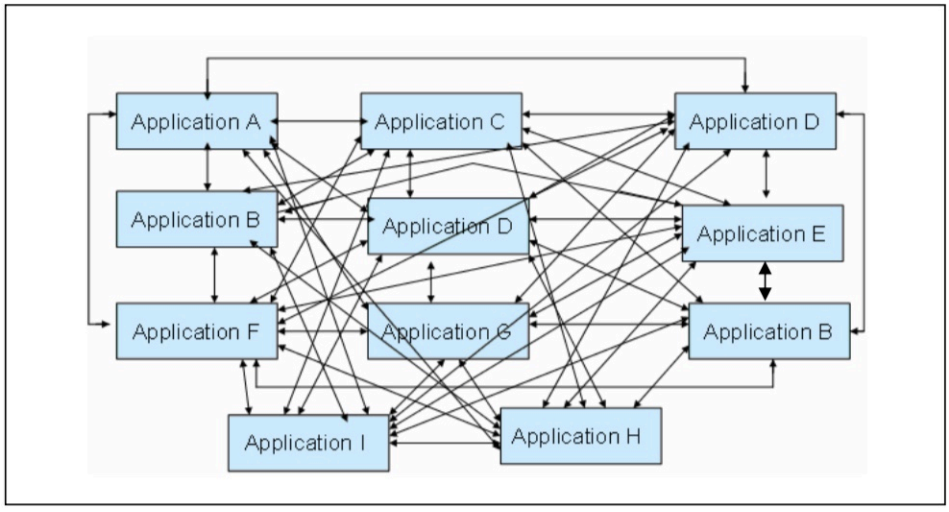


Fig 2 - Gartner Group – Syndrome Spaghetti [16]

De plus, ce type d'architecture pose de nombreux problèmes de reprise sur incident, coûts d'évolution et gestion de la complexité.

2.3.2) Les contraintes internes et externes du Système d'Information

Actuellement, les données échangées, souvent éparses et hétéroclites, représentent un volume extrêmement important qui tend à grossir au fil des années. Cela entraine de nombreux problèmes qui poussent les entreprises à urbaniser leur Système d'Information, décision motivée par deux types de contraintes, internes et externes. [8]

2.3.2.1) Les contraintes internes

L'apparition de nouveaux impératifs métier nécessite d'avoir de plus en plus l'information en temps réel (gestion des stocks, grille des vidéos visibles sur des sites de diffusion en différé, billetteries, etc.)

Par le phénomène de multiplication des systèmes et applications, les données opérationnelles de chaque système se retrouvent éparpillées, dupliquées, désynchronisées, incohérentes, ce qui a des répercussions sur la performance de l'entreprise :

- désaccords internes sur la valeur à attribuer à telle ou telle donnée
- faible contrôle des données
- coût financier pour réparer les erreurs liées aux données erronées
- altération des données partagées au fil des traitements par la chaine d'applications intermédiaires [3][10]
- augmentation des contraintes métier pour garantir leur cohérence, entraînant une explosion de leur complexité. Si une donnée est modifiée, comment en mesurer les impacts ? Les contraintes métier seront-elles toujours respectées pour chaque application ? [3]
- dysfonctionnement opérationnel dans les processus métier critiques

2.3.2.2) Les contraintes externes

Depuis 2005, on a assisté à une croissance de la complexité des contraintes réglementaires (Bâle II, IFRS, MIF 2, etc.), ce qui a contraint les entreprises à collecter davantage de données, à

justifier leurs opérations et à être plus transparentes au niveau de la présentation de leurs résultats. [8]

Parmi les contraintes légales figurent :

- la publication des comptes : un compte de résultat et un bilan doivent être produits annuellement par les entreprises. Celles cotées en bourse doivent également produire un tableau des flux de trésoreries
- le Système d'Information comptable : à partir de 2005, les sociétés européennes cotées doivent établir leurs états financiers trimestriellement selon le référentiel International Financial Reporting Standards (IFRS) [6][8]. Cette norme apporte davantage de transparence et son orientation est majoritairement tournée vers les investisseurs. Les logiciels de comptabilité se doivent de la respecter
- le respect de la règlementation de la Commission nationale de l'informatique et des libertés (CNIL) : un client peut demander des droits d'accès, de rectification ou de radiation d'une base de données d'une entreprise
- le Règlement général sur la protection des données (RGPD) : une nouvelle loi européenne, le RGPD, a pris effet le 25 mai 2018. Le RGPD donne aux individus résidant dans l'Union Européenne davantage de contrôle sur l'utilisation de leurs données et définit certaines obligations pour les entreprises qui traitent les informations de ces individus. Les logiciels utilisés doivent permettre de rectifier ou d'effacer des données inexactes (et pouvoir les supprimer définitivement) et garantir la sécurité des données. Une violation des données personnelles impose à l'employeur une notification à la CNIL et au client. [20][21]

2.4) L'enjeu de la performance

Un SI performant permet à une entreprise :

- d'améliorer sa productivité et réactivité grâce à l'assistance informatique (par l'optimisation des processus qui permettra de réduire les coûts)
- de remplacer des tâches récurrentes à faible valeur ajoutée par des traitements automatisés
- d'améliorer la relation client grâce à une vision unique et cohérente (et donc d'échanger des données informatisées qualitatives entre les différentes applications)
- de générer des indicateurs fiables pour aider à l'analyse et la prise de décisions

Un SI est jugé performant par :

- son niveau d'agilité : capacité à s'adapter aux changements métiers, légaux et/ou technologiques
- sa performance technique : temps de réponse, capacité à monter en charge, robustesse et fiabilité
- son architecture : centralisée, client/serveur ou orientée services (SOA)

La compétitivité, la mondialisation et l'ouverture des marchés ont poussé les entreprises à répondre à la pression des actionnaires, à vouloir anticiper les tendances du marché et mettre en place régulièrement de nouvelles offres pour répondre aux clients exigeants. L'organisation des entreprises est aujourd'hui centrée sur le client afin de lui proposer le bon produit au bon moment sur le bon canal. [8]

Face à des impératifs de rentabilité, les entreprises doivent pouvoir se baser sur le SI pour faire face à l'accroissement des risques opérationnels. Si les informations sur lesquelles elles se basent pour piloter l'entreprise (pour une analyse des ventes par région/département, par exemple) sont erronées ou obsolètes, l'impact sur leur performance s'en ressentira. [8]

Afin d'assurer la meilleure qualité possible des informations, il est important d'appliquer une stratégie de Gestion de Qualité des Données (« Data Quality Management » en anglais (DQM)) qui aura un impact direct sur la performance des entreprises.

3) La Gestion de la Qualité des Données

3.1) Définition

La DQM regroupe toutes les actions et procédures destinées à garantir et maintenir la qualité des données au sein d'une entreprise (nous allons ici nous cantonner aux données de référence).

La norme 9000 de l'Organisation Internationale de Normalisation (ISO) définit ainsi la qualité : « aptitude d'un ensemble de caractéristiques (3.10.1) intrinsèques d'un objet (3.6.1) à satisfaire des exigences (3.6.4) » [17]

Selon les auteurs F. Régnier-Pécastaing, M. Gabassi, et J. Finet, « la qualité des données est l'aptitude de l'ensemble des caractéristiques intrinsèques des données (unicité, exhaustivité,

fraîcheur, disponibilité, cohérence fonctionnelle, cohérence technique) à satisfaire des exigences internes (pilotage, prise de décision, etc.) et des exigences externes (réglementations par exemple) à l'organisation ». [2]

Pour y parvenir, il convient tout d'abord de se pencher sur les critères de qualité.

3.2) Les critères de qualité

3.2.1) Les critères intrinsèques

- l'unicité : la donnée doit n'être présente qu'une et une seule fois en base de données. Elle a un identifiant (sa clef primaire, par exemple, ou une clef composite comme le triplet nom/prénom/adresse) et ne doit pas être redondante.
- la complétude : concerne les schémas (tous les attributs à stocker existent-ils bien en base ?), les données (les données présentes ont-elles les champs requis remplis ?) et la population (la liste complète des données est-elle présente ?) [18]
- l'exactitude : une donnée est considérée exacte si elle est valide et précise (par rapport à des ordres de grandeur, des valeurs de référence ou issues d'un référentiel de confiance)
- la conformité : concerne le respect d'un ensemble de contraintes (un numéro de téléphone à 10 chiffres qui commence par un 0 ou +33, par exemple)
- la cohérence : une donnée peut être cohérente par rapport à une valeur de référence (par exemple le prix soldé d'un produit est forcément inférieur au prix du produit non soldé) ou par rapport à une autre donnée (un enfant a forcément une date de naissance postérieure à celle de ses parents)
- l'intégrité : se définit par les relations obligatoires d'une donnée par rapport une autre (par exemple, une facture doit être liée à une commande. Si elle ne l'est pas, la donnée « facture » a un problème d'intégrité)

3.2.2) Les critères de service

- l'actualité : la donnée doit toujours être la plus à jour possible. Si elle ne l'est plus à un instant T, on dit qu'elle est obsolète (puisque modifiée mais ne reflétant pas son nouvel état) ou que sa représentation a changé (elle peut être présente sous forme d'un tableau à un instant T et être à présent représentée sous forme d'un texte dont les valeurs sont séparées par des virgules)

- l'accessibilité : elle concerne la facilité d'accès aux données, c'est-à-dire que la donnée est à la fois accessible en mode requête à un service (service web, par exemple), en mode batch pour des synchronisations de masse (pour le décisionnel, par exemple), et en mode événement (présente à chaque mise à jour)
- la pertinence : définit l'utilité d'une donnée. Elle doit être accessible avec le niveau de granularité adéquat à son usage. [7]
- la compréhensibilité : la donnée est compréhensible s'il est facile de retrouver la bonne information parmi les attributs disponibles d'un objet peu importe le biais (processus, utilisateur, application)

3.2.3) Les critères de sécurité

- disponibilité : aptitude du système à remplir une fonction pour un service précis
- intégrité : concerne la légitimité de toute modification de donnée
- confidentialité : accès limité à qui de droit pour les besoins de service [7]
- traçabilité : conservation des opérations effectuées et de leurs auteurs

3.3) L'impact négatif sur la performance

Les risques les plus couramment rencontrés d'une faible qualité des données sont les suivants [6]:

- insatisfaction du client : une donnée non qualitative pourra empêcher de cibler un prospect dans une campagne marketing ou programmera l'envoi à un client d'un courriel relatif à une offre à laquelle il a déjà souscrit. Une adresse de livraison peut entraîner un retard ou un retour
- non-conformité des chiffres publiés : des erreurs dans les données comptables peuvent entraîner de lourdes amendes et peines
- dévalorisation de l'image de l'entreprise : des erreurs portant sur les données diffusées à l'extérieur de l'entreprise détériorent rapidement l'image de celle-ci et handicape son expansion commerciale. Si l'entreprise est cotée en bourse, l'impact boursier entraîne la chute rapide du cours de l'action
- perturbation du fonctionnement opérationnel : des données incorrectes ralentissent le fonctionnement opérationnel et rendent les décisions difficiles ou incertaines ; cela peut aussi générer du stress et de l'insatisfaction chez les employés ; une perte de confiance

dans les données s'installera, et un surcroit de travail engendré par les erreurs de données sera inévitable

- erreurs de stratégie : définir une stratégie devient difficile et chronophage si les informations utilisées sont de faible qualité ou obsolètes
- augmentation des coûts : le coût lié au retraitement des données et à la perte de clients peut représenter jusqu'à 10% du chiffre d'affaires, ce qui ampute de façon importante la rentabilité

Les problèmes les plus courants liés à la qualité sont :

- les données de départ : si les données intégrées sont peu qualitatives dès le départ (provenant d'un fichier excel non contrôlé par exemple), on assistera au phénomène du « Garbage In, Garbage Out » (GIGO), c'est-à-dire que si les données de départ sont mauvaises, les sorties le seront également
- les doublons : En cas de doublons mal maîtrisés, quel enregistrement du client ou du fournisseur l'employé doit-il sélectionner ? Lors d'une mise à jour, certains enregistrements seront corrects tandis que les doublons ne le seront pas (ou chaque version sera partiellement correcte). Les analyses par client deviennent difficiles. Le cycle de vie des données est perturbé si un client invoque le droit à l'oubli dans le cadre de la RGPD
- les coordonnées des partenaires : l'expédition des commandes à la mauvaise adresse n'est pas si rare. Ces erreurs engendrent des difficultés et de la perte de temps pour communiquer. Une connaissance affaiblie des partenaires entrainera une baisse de la réactivité et des affaires en général
- les indicateurs d'activité : le chiffre d'affaires, le coût de revient des produits, la marge, le carnet des commandes etc. La qualité des résultats fournis par les systèmes décisionnels dépend directement de la qualité des données qui leur sont fourni [2]. Si elles sont basées sur des informations peu qualitatives, les analyses du management seront incorrectes et les décisions résultantes inappropriées

3.4) L'impact positif sur la performance

Une bonne qualité des données aura un impact significatif sur les points suivants :

- réactivité : si un client doit recevoir sa commande pour le soir même, la justesse de son adresse postale augmentera considérablement les chances de livraison dans les délais impartis [7]

- connaissance des clients : mieux connaître le marché et ses clients permet de tisser des liens forts et durables et garantit la pérennité de l'entreprise. Avec les données clefs permettant de correctement communiquer avec le client (courriel, téléphone, adresse, fax, etc.), les opérations marketing sont fluidifiées
- gestion opérationnelle optimisée : si les données manipulées sont fiables, elles sont perçues par les employés comme un allié leur permettant d'optimiser leur travail et la performance globale de l'entreprise. Le temps passé à corriger les données est largement diminué et rentabilisé pour pouvoir appeler davantage de clients, par exemple
- pertinence des décisions : coût de revient (a) et clients stratégiques (b)

(a) Concernant la pertinence des décisions, dans le cas du calcul du coût de revient d'un produit fabriqué, si un logiciel se base sur des informations exactes et précises, le résultat sera fiable. Il n'y aura pas de risque d'arrêter à tort la production d'un produit rentable ou de produire davantage un produit trop coûteux.

(b) De même, si la base de données est saine, on ne risquera pas de faire l'impasse sur certains clients stratégiques qui auraient pu être enregistrés plusieurs fois et n'apparaitraient pas de façon consolidée comme des clients importants.

3.5) Un besoin d'urbanisation du Système d'Information

Selon l'auteur E. Abou-Chakra, l'urbanisation se définit comme étant « l'ensemble des moyens permettant de faire évoluer le Système d'information et le Système informatique aux mêmes rythmes que la stratégie et l'organisation. Elle consiste à décrire la structuration du système cible et la façon de l'atteindre. » [11]

Les contraintes du SI ont fait naître un véritable besoin de son urbanisation avec notamment la mise en place de données uniformisées, uniques, pérennes, fiables, centralisées et synchronisées sur lesquelles les systèmes opérationnel et décisionnel se baseront : les données de référence.

4) Les données de référence

4.1) Définition et typologie

Une entreprise de e-commerce, par exemple, a plusieurs processus métier tels que [2] :

- la vente, qui repose sur des données « *produits* » et « *clients* »
- la facturation et la comptabilité d'entreprise qui s'appuient sur des données telles que les *prix* et les *comptes*
- le pilotage de l'ensemble qui effectue la consolidation de toutes les données mobilisées lors des transactions

Tous les processus métier utilisent des données comme éléments de base : unités, devises, modes de paiement [6], clients, adresses, fournisseurs, employés, articles, contrats, offres [2], taux, pays, unités de mesure, filiales, départements, branches, codes postaux, configurations et paramètres [8], etc.

Ces éléments sont déclarés (et non créés par un processus transactionnel comme le processus de vente qui crée les commandes et les factures qui sont des données dites opérationnelles).
On appelle ces données des « données de référence » ou « Master Data » en anglais.

Elles sont donc partagées par l'ensemble des processus qui soutiennent l'activité courante d'une entreprise et ses prises de décision [2] et sont utilisées tant par les applications opérationnelles (comme la vente ou la comptabilité) que par les applications de support (tels que les créations et productions de rapports).

Du point de vue du responsable SI, une donnée de référence est une donnée du SI, partagée et utilisée sans modifications par ses applications. En effet, la création ou la modification d'une donnée de référence est réalisée au travers de « processus référentiel métier ». [2]

La donnée de référence doit être de qualité et respecter des règles communes à l'ensemble des entités utilisatrices. [2]

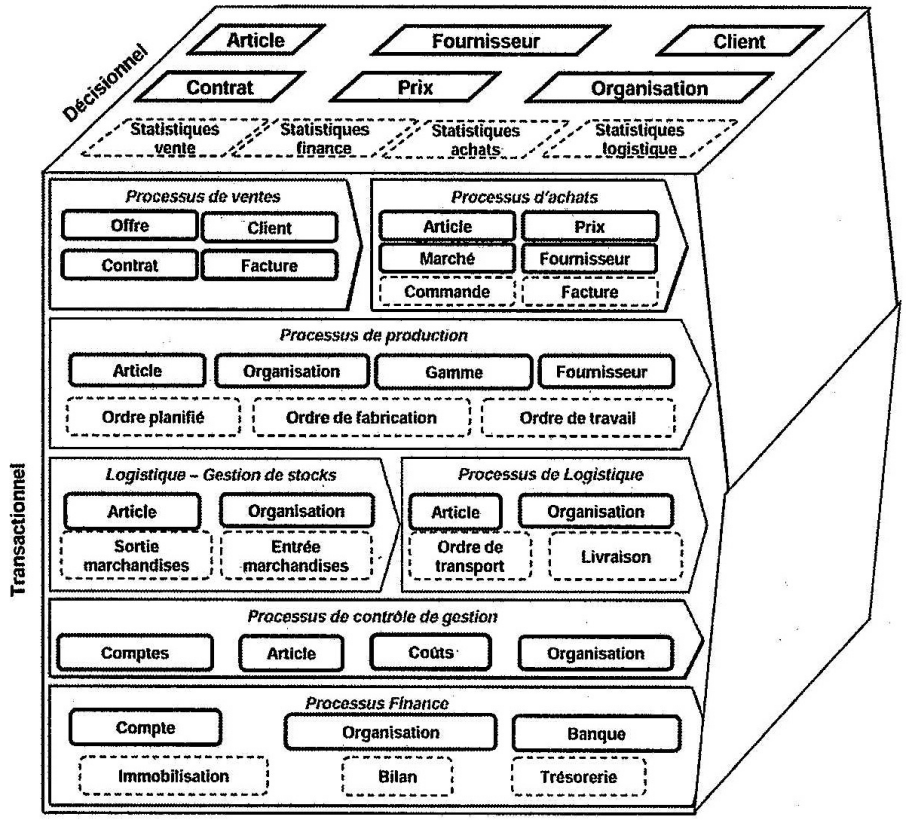


Fig 3 - Quelques données de référence dans un Système d'Information [2]

4.2) Dictionnaires de données

Les données de références sont décrites dans des dictionnaires de données [2][9] qui sont des collections de métadonnées de trois types : métier (aussi appelées descriptives), techniques (aussi appelées structurelles) et administratives. [2][12]

Les métadonnées, littéralement « données sur la donnée », sont des descriptions métier et/ou technique d'une donnée. [2] Elles servent de vocabulaire commun à l'organisation et sont un outil de dialogue entre la Maîtrise D'Œuvre (MOE) et la Maîtrise d'Ouvrage (MOA).

4.2.1) Métadonnées métier

Les métadonnées métier (« business metadata » en anglais) définissent une correspondance entre les objets métier et les objets techniques (c'est-à-dire existant en base de données).
Elles se présentent sous forme d'un tableau de cinq colonnes [13]: nom d'un attribut d'un objet métier, code du champ en base de données, type de données (numérique, texte, caractères alphabétiques, etc.), longueur/dimension et commentaire.

Ainsi, pour l'objet métier Client, par exemple, le tableau pourra se présenter de la sorte :

Tableau I : extrait de dictionnaire de métadonnées métier

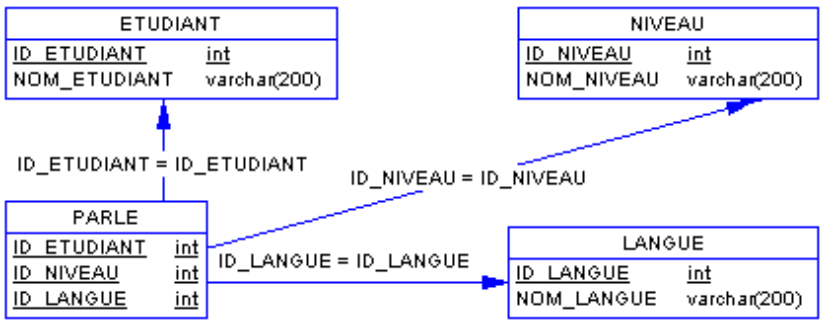
Client : table customer				
nom	code	type	Longueur/dimension	commentaire
Nom	lastname	TEXTE	64	
Prénom	firstname	TEXTE	64	
Code postal	zipcode	TEXTE	10	Est alphanumérique pour accepter les codes postaux de la Corse
Adresse	address	NOMBRE	10	Fait référence à l'identifiant de la table address_customer

Ces métadonnées permettent aux utilisateurs de connaître la nature (type de chaque champs) et l'origine (la/les tables concernées en base de données) dans leurs rapports. Elles permettent également aux développeurs et utilisateurs de mieux comprendre les données de référence afin de définir et parfaire la construction des mêmes rapports en communiquant.

4.2.2) Métadonnées techniques

Aussi appelées métadonnées structurelles (« technical metadata » en anglais), elles fournissent les informations sur la structure interne des tables de la base de données afin d'en construire les modèles de données telles que les Modèles Physiques de Données.[14]

Tableau II : Modèle Physique de Données avec liaisons de clefs étrangères



Tout ce qui a trait aux clefs primaires & étrangères, triggers, index, permissions sont définies dans ces métadonnées.[15]

Elles servent à piloter des outils comme Extract Transform Load (ETL, technique servant à récupérer les informations de différentes sources pour les nettoyer/transformer en vue de les stocker dans un entrepôt de données qui permettra au système décisionnel de créer des indicateurs d'analyse et d'aide à la décision.)

Ces métadonnées permettent également :

- d'analyser le code et les bases de fichiers
- de tracer l'origine de données calculées
- de spécifier, maîtriser et maintenir les référentiels des outils d'aide à la décision (Business Intelligence), les applications pour les alimenter, les calculs dans les rapports, la qualité des données
- d'effectuer les analyses d'impact lors des modifications et évolutions des systèmes décisionnels (« Business Intelligence » en anglais) et de leurs sources

4.2.3) Métadonnées administratives

Aussi appelées métadonnées opérationnelles (« operational metadata » en anglais), elles servent à :

- savoir qui a modifié la donnée (auteur, historique, ordonnancement des modifications)
- connaître les différentes transformations effectuées sur la donnée
- savoir si la donnée est active/supprimée/archivée

4.3) Nécessité d'une gestion des données de référence

Toutes ces données de référence doivent être gérées afin d'en garantir l'unicité, la fraicheur, la pérennité, la qualité, l'intégrité et la complétude. Une architecture de centralisation doit être choisie, un référentiel de données mis en place, une stratégie de réharmonisation, de dédoublonnage et de consolidation mise en place. Tout cela se fait par la Gestion des Données de Référence (GDR), encore appelé « Master Data Management » (MDM) qui jouera un rôle dans la performance du SI.

5) Gestion des Données de Référence

5.1) Définition et évolution du périmètre des données traitées

Le MDM est une discipline des technologies de l'information concernant les données de référence. Elle consiste à les centraliser, à garantir leur qualité, cohérence et partage entre diverses architectures de systèmes et fonctions métier.

Elle apparait en 2005 et concerne à l'origine les données de référence des produits et des clients [3]:
- la gestion des catalogues de produits (« Product Information Management » en anglais) notamment dans les domaines de la grande distribution et du manufacturing,
- l'intégration des données de clients (« Customer Data Integration » en anglais) particulièrement pour l'administration de grosses bases de données transactionnelles (gestion des doublons, vérification et homogénéisation des adresses, etc.) dans des domaines tels que les banques ou les assurances.

Elle s'est aujourd'hui généralisée à d'autres domaines (cf 4.1 Définition et Typologie des données de référence) dont la finance pour l'établissement des rapports de gestion et des comptes annuels et consolidés. [8]

5.2) Fonctions principales assurées par l'architecture Master Data Management

Le MDM se base sur trois approches fondamentales que sont l'intégration des données de référence, leur gouvernance et leur qualité.

5.2.1) L'intégration des données de référence

Les données de référence sont gérées par un référentiel de données. Selon le Cadre Commun d'Architecture des Référentiels de données du gouvernement, « un référentiel de données est un dispositif permettant la gestion mutualisée et le pilotage dans le temps, de données de référence, à savoir, l'actualisation et la mise à disposition de ces données. Un référentiel de donnée comprend aussi les éléments permettant de piloter et faire évoluer cette gestion de données de référence, en fonction des besoins des « clients » du référentiel, et en fonction des problèmes de qualité de données rencontrés. » [1]

La mise en œuvre d'un référentiel vise à créer un point focal au seuil du SI, appelé « point de vérité » [2]. Il garantit la validité des informations détenues par le référentiel à partir de ce point. On appelle « points d'acquisition » les points d'entrée d'une donnée au sein du SI.

Les architectures classiques pouvant être mise en place sont des architectures de [2] :

- consolidation : plusieurs sources de données (points d'acquisition) alimentent le référentiel avec leurs données propres. Ce dernier joue le rôle de consolidateur et met la donnée à disposition des applications consommatrices. Le DQM est requis si la volumétrie est importante.
- coopération : les applications existantes sont utilisées comme points d'acquisition de la donnée. Elles utilisent les données du référentiel et partagent les processus de création et de modifications des données avec lui. Toute saisie est fonctionnellement validée par le référentiel avant de pouvoir être utilisée. Le référentiel peut soit agréger les données des applications sources, soit recenser les attributs positionnés dans les applications source (système d'enregistrement distribué). La mise à jour du SI se fait par événements déclenchés par le référentiel après mise à jour de sa donnée.
- centralisation : points d'acquisition et de vérité fusionnent. Les processus référentiels de création/modification des données sont gérés au niveau du référentiel. Les applications en aval consomment les attributs de la donnée référentielle. Cette architecture autorise une gouvernance renforcée.

5.2.2) La gouvernance des données de référence

Selon le MDM Institute, la gouvernance des données se définit par « l'orchestration formelle des personnes, des processus et de la technologie permettant à une organisation de tirer parti des données en tant qu'actif de l'entreprise. » [15]. Elle définit une collection de bonnes pratiques et de moyens facilitant la gestion des données [3] et concerne l'ensemble des droits et règles affectées à chaque utilisateur. [2][16]

La gouvernance des données sous-entend quatre actions [3] :

- prévoir : établir une stratégie de gestion des données déterminant les procédures afin de mettre à jour, partager, assurer la sécurité, contrôler l'accessibilité des données, préserver les droits de leur(s) propriétaire(s), etc.

- négocier : trouver des compromis entre les fournisseurs et les consommateurs de données tout en prenant en compte les réglementations en vigueur
- soulever des problèmes : être capable de détecter des anomalies relatives aux données et à leur utilisation
- résoudre des problèmes : être capable de prendre les mesures nécessaires pour corriger ces anomalies ou en diminuer au maximum les effets négatifs.

Les principaux moyens préconisés sont [3] :

- la création d'un comité de pilotage et de surveillance
- l'identification des « propriétaires », des « fournisseurs » et des « consommateurs » des données
- la définition des rôles et responsabilités
- la description des données (glossaire métier, dictionnaires de métadonnées)
- l'établissement des normes et des procédures pour l'utilisation des données
- la mise en œuvre des vérifications et des contrôles par des indicateurs de qualité des données et des outils d'auditabilité des données [2]

Les données étant partagées entre différentes applications et intervenants, il est primordial de mettre en place des règles d'accès et de modification. La sécurité des données sera analysée à de multiples niveaux : attribut, objet, référentiel, processus (possibilité par exemple de consulter des données mais pas de les modifier). Il conviendra de distinguer différents niveaux de confidentialité, d'identifier les besoins de traçabilité, de définir les droits d'accès…
En effet, si la donnée était modifiable par des personnes dont ce n'était pas le rôle et qui n'étaient pas à même de d'appréhender la portée des données modifiées, la qualité des données s'en trouverait grandement affectée.

La gouvernance a également pour objectif l'enrichissement du spectre informationnel porté par les données ou corrélé grâce à elles (Enterprise Information Management (EIM)). [2]

5.2.3) La qualité des données de référence

Les données de référence, comme nous l'avons vu précédemment, peuvent être dispersées dans plusieurs bases de données. Le taux d'anomalies augmente fortement en fonction du nombre de sources de données et du degré d'hétérogénéité de celles-ci.

Dans le MDM, le cycle de vie de ces données ressort comme étant une notion fonctionnelle importante. Il est en effet primordial de travailler avec les métiers pour reconstituer et cartographier le cycle de vie des données, en connaître l'origine et quel sera leur usage [10], sous peine de voir leur qualité se dégrader au fil du temps.

Le cycle de vie se définit autour de fonctions telles que [8] :

- la découverte et le profilage des données
- l'acquisition et l'intégration des données
- la maintenance des données
- l'usage des données
- l'archivage et la destruction des données

L'usage des données est une notion essentielle en qualité de données. Elle se définit par rapport aux traitements métiers de ces données. C'est alors que l'on dit que la qualité est appropriée à l'usage qu'on en fait (« fitness for use »). [8]

Également, il sera très important que l'outil MDM retenu gère les accès concurrents (lorsque deux programmes tentent de modifier une donnée en même temps, la donnée modifiée par le premier programme doit être répercutée sur le second) lors des mises à jour des données afin d'en conserver la cohérence.

Le contrôle et l'amélioration de la qualité des données sont des éléments primordiaux à toute approche MDM. De plus, la centralisation des données, leur dédoublonnage, le contrôle de leur intégrité et cohérence (par le DQM) permettront aux entreprises appliquant le MDM de rapidement réagir, notamment en ce qui concerne les obligations légales dans le cadre de la RGPD. [20]

De ce fait, les processus transactionnels se feront se manière optimale, et le système décisionnel se basera sur les indicateurs de pilotage de qualité, ce qui aura un impact significatif sur la performance de l'entreprise.

Conclusion

De 1960 à nos jours, la donnée, toujours plus complexe, spécialisée et volumineuse, a vu son importance grandir au sein du Système d'Information des entreprises. Soumis à des contraintes internes et externes, ce dernier se base sur des données provenant de sources différentes afin d'alimenter des applications, générer des documents financiers, des rapports d'activité et créer des indicateurs de pilotage de performance et des activités de l'entreprise.

Ces indicateurs, générés par le système décisionnel (encore appelé « Business Intelligence ») faisant partie du Système d'Information, pour être fiables, ont besoin de données qualitatives issues des systèmes opérationnels afin de définir une stratégie d'entreprise et prendre les décisions adéquates par rapport au marché.

Or, ces données, concernant les clients, produits, prix, gammes etc., souvent dispersées dans plusieurs bases de données ou fichiers, peuvent se révéler obsolètes, incohérentes, dupliquées et donc inexploitables. Les conséquences pour l'entreprise peuvent être catastrophiques : pertes de clients, perte de temps et coût lié au retraitement des données erronées, documents financiers légaux erronés, erreurs stratégiques, etc.

Partagées entre les différents services du Système d'Information, ces données, dites de référence, doivent être unifiées, décrites, contrôlées, tracées et synchronisées afin d'en assurer la pérennité. Grâce au Master Data Management, elles seront centralisées et partagées dans toute l'entreprise par de nombreuses applications. Le Master Data Management permettra également aux entreprises de répondre de façon plus efficace aux exigences de la RGPD.

La qualité de ces données, assurée par le Master Data Management, jouera alors un rôle décisif dans la performance de l'entreprise concernant sa réactivité, la connaissance des clients, la pertinence de ses décisions et l'optimisation du travail des employés.

Le Master Data Management s'inscrit dans une logique d'urbanisation du Système d'Information qui poussera l'architecte à réfléchir aux échanges des données et à restructurer, parfois sur plusieurs années, tout le Système d'Information de l'entreprise.

Le Master Data Management
et son rôle dans la performance d'une entreprise

Mémoire présenté en vue d'obtenir
UE « Information et Communication pour Ingénieur »
Spécialité : INFORMATIQUE
C.N.A.M., Paris 2018

RESUME

Ces dernières années, les données ont eu tendance à devenir toujours plus complexes, spécialisées et volumineuses. Au cœur du Système d'Information, elles alimentent les applications qui, par des processus transactionnels, permettent au système décisionnel de générer des indicateurs de pilotage afin de définir une stratégie d'entreprise.

Les données essentielles, dites de référence, concernent généralement les clients, fournisseurs, produits, etc. Souvent éparses, désynchronisées et incohérentes, elles sont alors un frein au bon fonctionnement de l'entreprise et affectent sa performance.

La Gestion des Données de Référence s'assurera qu'elles soient centralisées, partagées, contrôlées et de qualité. Cela jouera un rôle décisif dans la performance de l'entreprise concernant sa réactivité, la connaissance de ses clients, la pertinence de ses décisions et l'optimisation du travail de ses employés.

Mots clés : Système d'Information, système décisionnel, Gestion des Données de Référence, qualité, performance

---

SUMMARY

Nowadays, data has tended to become more complex, specialized and higher volume. At the center of Information Systems, data feeds applications by using transactional processes. This enables business intelligence to generate operational indicators that define business strategy.

Essential data, so called master data, is generally about customers, suppliers, products, etc. Often sparse, out of sync and inconsistent, such data prevents the proper functioning of the company and affects its performance.

Master Data Management will ensure the centralization, sharing, control and quality of the master data. This will play a decisive role in the company's performance regarding its responsiveness, customer knowledge, relevance of its decisions and optimization of the work of its employees.

Keywords : Information System, Business Intelligence, Master Data Management, quality, business performance

Bibliographie

Chapitres dans un ouvrage imprimé

[2] F. Régnier-Pécastaing, M. Gabassi, et J. Finet, *MDM enjeux et méthodes de la gestion des données*. Chapitres 1, 2, 9. Paris: Dunod, 2008.
[5] R. Reix, B. Fallery, et M. Kalika, *Systèmes d'information et management*, 7e édition. Chapitre 1. 2016.
[6] C. Brasseur, *Data management qualité des données et compétitivité*. Chapitres 1, 3, 4. Paris: Hermes science publ. Lavoisier, 2005.
[7] J. Bizingre, J. Paumier, et P. Rivière, *Les référentiels du système d'information données de référence et architecture d'entreprise*. Chapitres 1, 4. 2013.
[9] J.-F. Pillou et P. Caillerez, *Tout sur les systèmes d'information grandes, moyennes et petites entreprises*, 2e édition [mise à jour]. Chapitres 5, 6. Paris : Dunod, 2011.

Travaux universitaires

[16] KAMALAK Axel. Les apports de la méthode MDM dans la performance du SI des entreprises. Mémoire de fin d'étude, École Privée des Sciences Informatiques, 105 p, 2008. Disponible sur : http://docplayer.fr/7940950-Les-apports-de-la-methode-mdm-dans-la-performance-du-si-des-entreprises.html (consulté le 23 avril 2018)

Sites web

[1] L'Etat. Urbanisation du système d'information de l'Etat | Les documents de référence du S.I. de l'État, (en ligne). Disponible sur : http://references.modernisation.gouv.fr/urbanisation-du-systeme-dinformation-de-letat (consulté le 25 avril 2018).
[3] J.-C. Trigaux, Master Data Management, (en ligne). Disponible sur : https://www.smalsresearch.be/download/research_reports/deliverable/MDMReport_FR_Final_29-12-09.pdf (consulté le 23 avril 2018)
[4] LeMagIT, Que signifie Données semi-structurées? - Définition par WhatIs.com, (en ligne). Disponible sur : https://www.lemagit.fr/definition/Donnees-semi-structurees (consulté le 25 avril 2018).
[8] CNRS, L. Dubrovin, Gestion des données de référence (MDM), (en ligne). Disponible sur : http://www.dsi.cnrs.fr/description_referentiel/donnees/Documents/Synthese_CXP.pdf (consulté le 23 avril 2018)
[10] MARIKO Dominique, Intd CNAM, MDM _ DQM 2016-1, (en ligne). Disponible sur : https://www.scribd.com/document/364810494/MDM-DQM-2016-1 (consulté le 23 avril 2018).
[11] Éric Abou-Chakra, CNAM, Urbanisation & Architecture des Systèmes d Information, (en ligne). Disponible sur : http://docplayer.fr/20239037-U-arsi-urbanisation-architecture-des-systemes-d-information-eac-op-et-jcc.html (consulté le 26 avril 2018).
[12] Cornell University Library, Table 5.1 Metadata Types, (en ligne). Disponible sur : http://preservationtutorial.library.cornell.edu/metadata/table5-1.html (consulté le 27 avril 2018).

[13] EWSolutions, Metadata Management Fundamentals, (en ligne). Disponible sur : https://www.ewsolutions.com/metadata-management-fundamentals/ (consulté le 23 avril 2018).
[14] DATAVERSITY, LDM Webinar: Data Modeling & Metadata Management, (en ligne). Disponible sur : https://www.slideshare.net/Dataversity/data-modeling-metadata-management (consulté le 23 avril 2018)
[15] MDM Institute, What is Data Governance?, (en ligne). Disponible sur : http://0046c64.netsolhost.com/whatIsDataGovernance.html (consulté le 28-avr-2018).
[17] Organisation Internationale de Normalisation, « ISO 9000:2015(fr), Systèmes de management de la qualité — Principes essentiels et vocabulaire », (en ligne). Disponible sur : https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v2:fr (consulté le 29 avril 2018).
[18] EXQI, Les dimensions de la qualité des données, (en ligne). Disponible sur : http://exqi.asso.fr/site/medias/641739FicheDimensionsQualite.pdf (consulté le 29 avril 2018).
[19] Ben El Arbi SABBAR, La performance de l'entreprise, (en ligne). Disponible sur : http://sabbar.fr/management/la-performance-de-lentreprise (consulté le 29 avril 2018).
[20] Les echos, La GDPR comme tremplin vers une gouvernance des données gagnant-gagnant, (en ligne). Disponible sur : https://www.lesechos.fr/idees-debats/cercle/cercle-164098-la-gdpr-comme-tremplin-vers-une-gouvernance-des-donnees-gagnant-gagnant-2052069.php (consulté le 6 mai 2018).
[21] Gildas Neger, La gouvernance des données pour les nuls (RGPD), (en ligne). Disponible sur : https://www.village-justice.com/articles/gouvernance-des-donnees-pour-les-nuls-rgpd,26965.html (consulté le 6 mai 2018)

Liste des figures

Figure 1 : Le Système d'Information ……..…… 10
Figure 2 : Syndrome Spaghetti – Gartner Group …… 12
Figure 3 : Quelques données de référence dans un Système d'Information …… 21

Liste des tableaux

Tableau I : Extrait de dictionnaire de métadonnées métier …… 22
Tableau II : Modèle Physique de Données avec liaisons de clefs étrangères ……..…… 22