CDO 2020

Etat de l'art des centres d'intérêt du Chief Data Officer

Jérôme Capirossi

CDO 2020

L'état de l'art des centres d'intérêt du Chief Data Officer

Jérôme CAPIROSSI

UNEXX Editions - 2020

Ce livre a été publié par UNEXX, contact@unexx.eu

ISBN : …….

Table des matières

Avertissement

Cet ouvrage est tiré du séminaire : « Gouvernance et Architecture des données » d'UNEXX.

Il s'adresse naturellement aux Chief Data Officers, mais aussi plus largement à tous les acteurs de la data : les Data stewards, Data managers, les architectes Data, les Directeurs de projets data, Chefs de projet, Les Responsables d'applications,…

1 Introduction du sujet

Nous n'évoluons pas dans un jardin d'Eden des données où une Nature bienveillante les mettrait à disposition des entreprises qui n'auraient qu'à les collecter.

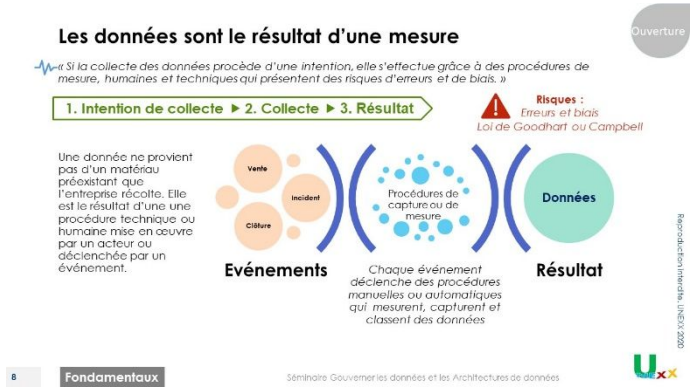
Les données sont le produit d'opérations de mesure qui requièrent une compréhension préalable, par les entreprises, de leur environnement.

Le fonctionnement régulier des systèmes d'information génère des orages de données qui par manque de connaissance ou par mauvaise qualité sont peu utilisées.

1.1 Les données sont le résultat d'une mesure

Les données ne se sont pas des objets qui préexistent dans la Nature et qu'il suffirait de récolter. Elles sont le résultat d'opérations de mesure dans le but de fournir des informations à propos d'un événement entrant dans le domaine d'intérêt de l'entreprise. Ce peut être, par exemple, un événement qui indique qu'un prospect recherche un produit ou un service que fournit l'entreprise.

La simple opération d'enregistrer l'identité d'un client doit aussi être considérée comme une mesure, excepté que, dans ce cas, la procédure est établie par l'état et le résultat figure sur les papiers d'identité du client. Nous verrons dans un chapitre ultérieur que cette donnée qui semble pourtant simple à acquérir, pose de nombreux problèmes pour la « gestion de la relation client ».



Une fois mesurée, la donnée entre dans le calcul d'un indicateur qui caractérise le comportement de l'objet observé. C'est l'interprétation de cet indicateur, souvent par comparaison à des seuils ou des normes, qui produit réellement l'information utilisée.

Cette chaine de valeur de la donnée ne peut être construite que si l'on a préalablement une idée du comportement de l'objet et des indicateurs à calculer. Cela n'est possible que grâce à la modélisation. Ainsi, on peut dire que toute donnée entre dans

une chaine de valeur qui est organisée selon l'idée a-priori que l'on a de la chose à observer.

Par exemple, si je crois qu'un prospect qui revient une 3ème fois sur mon site de e-commerce est intéressé par un article mais doit absolument bénéficier d'une remise pour compléter son acte d'achat, je vais suivre l'indicateur des retours des prospects sur mon site et lui appliquer la valeur seuil, 3.

Mais il se peut que plusieurs prospects ne reviennent jamais 3 fois, et abandonne avant, ou bien que certains prospects auraient quand même acheté le produit au prix standard lors de la 3ème visite.

Cet a priori lié à la mesure entraine des biais qui peuvent aussi poser des problèmes éthiques, s'ils incluent des considérations de race ou de croyance. D'autre part, la mesure agit toujours sur le système observé. Ainsi, les ristournes accordées lors des re-visites de mon site e-commerce ont été repérées par les internautes qui ont changé leur comportement afin de bénéficier de ces ristournes, ce qui n'était pas l'intention initiale.[1]

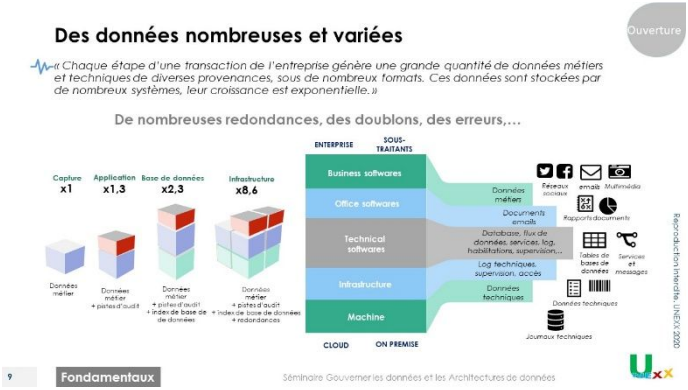
1.2 Des données nombreuses et variées

Les systèmes d'information sont constitués d'une superposition de couches (application métier, logiciels de base, infrastructures, machine) qui sont toutes mobilisées pour exécuter un traitement. Lorsqu'un utilisateur effectue une transaction avec son progiciel métier, ou lorsqu'il utilise son environnement bureautique, il déclenche une cascade d'opérations techniques qui se traduisent en données techniques qui finissent enregistrées dans les systèmes de stockage des infrastructures.

Chaque couche rend compte, dans un ou plusieurs journaux, de la réussite et du contexte d'exécution de l'opération qu'il a réalisée. Il doit également produire et enregistrer les données techniques requises par la technologie pour réaliser les fonctions

[1] Effet GoodHart ou Campbell

techniques qui lui reviennent. Par exemple, les gestionnaires de base de données doivent créer ou mettre à jour une entrée d'index pour chaque donnée métier.



Enfin, les infrastructures répliquent les données (sauvegarde, mirroring…), pour être, à tout moment, capables de résister à une panne et reprendre les opérations le plus rapidement possible.

En mettant bout à bout ces traitement, on peut affirmer que chaque opération métier génère un orage de données dans le système d'information.

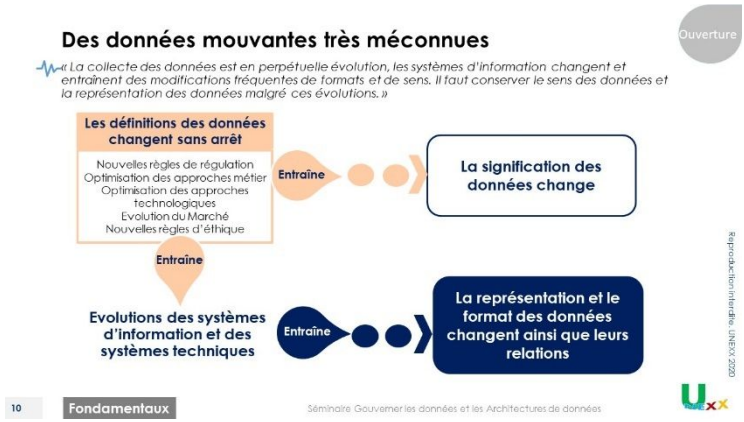
1.3 Des données mouvantes très méconnues

Les composants applicatifs qui forment les couches du système d'information connaissent de nombreuses évolutions qui font évoluer la définition et la structure des données.

Ce peut être un nouveau module progiciel qui créé un nouvel objet métier, ce peut être aussi une nouvelle fonctionnalité qui change l'utilisation d'une ou plusieurs données applicatives.

Par exemple, l'utilisation de la donnée TVA pour gérer une fonctionnalité de commissionnement d'un partenaire métier. Si ce n'est pas une bonne pratique, cela arrive lorsqu'on considère une facilité de réalisation locale, sans prendre en compte la

complexité d'évolution des applications avals comme le décisionnel.



Il résulte de cela que la structure des données et leur sens évoluent sans cesse.

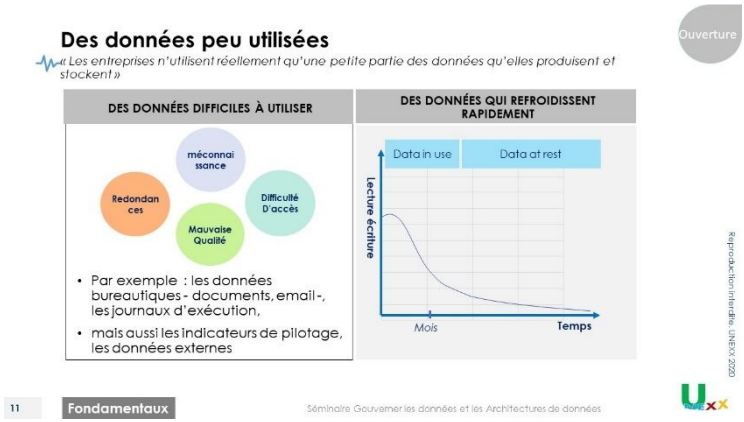
L'oubli de ces conventions avec le temps, ou l'absence de processus de notification d'évolution de la structure des données finit par rendre la documentation des données obsolète.

1.4 Des données peu utilisées

Les entreprises n'utilisent réellement qu'une partie des données qu'elles produisent.

- **Parce qu'elles ne les connaissent pas**. La documentation incomplète des progiciels ou des données techniques difficiles à comprendre entravent l'exploitation des données. Par exemple, il a fallu du temps et de l'outillage pour démocratiser l'usage des journaux techniques des serveurs http pour collecter et calculer des indicateurs de fréquentation de sites internet.
- **Parce que l'accès aux données est difficile.** L'utilisation de logiciels spécialisés pour accéder aux données ou bien la volonté de certaines directions de ne pas partager les

données, pour par exemple, des motifs de confidentialité, ne facilitent pas l'accès aux données.



- **Parce que les données sont de mauvaise qualité.** L'obligation de mettre en qualité les données avant leur utilisation alourdit les usages métiers. Cela concerne notamment les usages décisionnels avals ou les usages postérieurs à la vente ou à la production.
- **Parce que les données sont redondantes.** Lorsqu'une donnée est présente à plusieurs endroits du système d'information, on ne sait pas identifier la donnée fiable, à jour qui doit être prise en compte. Cela peut avoir des conséquences graves notamment dans certains usages de santé.

On constate que les données sont très utilisées lors des transactions de vente ou lors des activités de production, ce sont les « données chaudes ». Une fois reversées dans les historiques ou les entrepôts, elles sont souvent agrégées dans des indicateurs et puis conservées sans être sollicitées plus avant. Ce sont les « données froides ».

Le Big data en travaillant plus fréquemment les données de base (« données brutes ») change légèrement ce constat.

2 Valoriser les données de l'entreprise

Les données de l'entreprise ont de la valeur. La révolution du secteur des services n'a pu avoir lieu que grâce à une automatisation accrue des processus métier, rendue possible par une meilleure gestion des données.

A l'heure du Big Data et de l'intelligence artificielle, on comprend toute la valeur que prennent les données pour les nombreux usages qui en découlent.

Les grandes tendances d'évolution, la personnalisation, la logistique intégrée, l'industrie 4.0… ne pourront se réaliser sans une gestion des données performante.

Pour toutes ces raisons, aujourd'hui les données sont un aspect incontournable de la Stratégie d'entreprise.