

QuaRES,

pour la

Qualité

en Recherche

et en

Enseignement Supérieur

QuaRES

Association pour la Qualité
en Recherche et en Enseignement Supérieur

www.quares.fr

ATELIER : Maîtrise de l'information

Thierry BONTEMS
Responsable qualité
UMR PACTE - CNRS



Réfléchir ensemble à la notion de données et d'information

Trouver 10 causes à la mauvaise qualité des données

Trouver des solutions pour répondre à cette problématique

Un constat

Préhistoire

Petit territoire
Famille

Antiquité et Moyen Âge

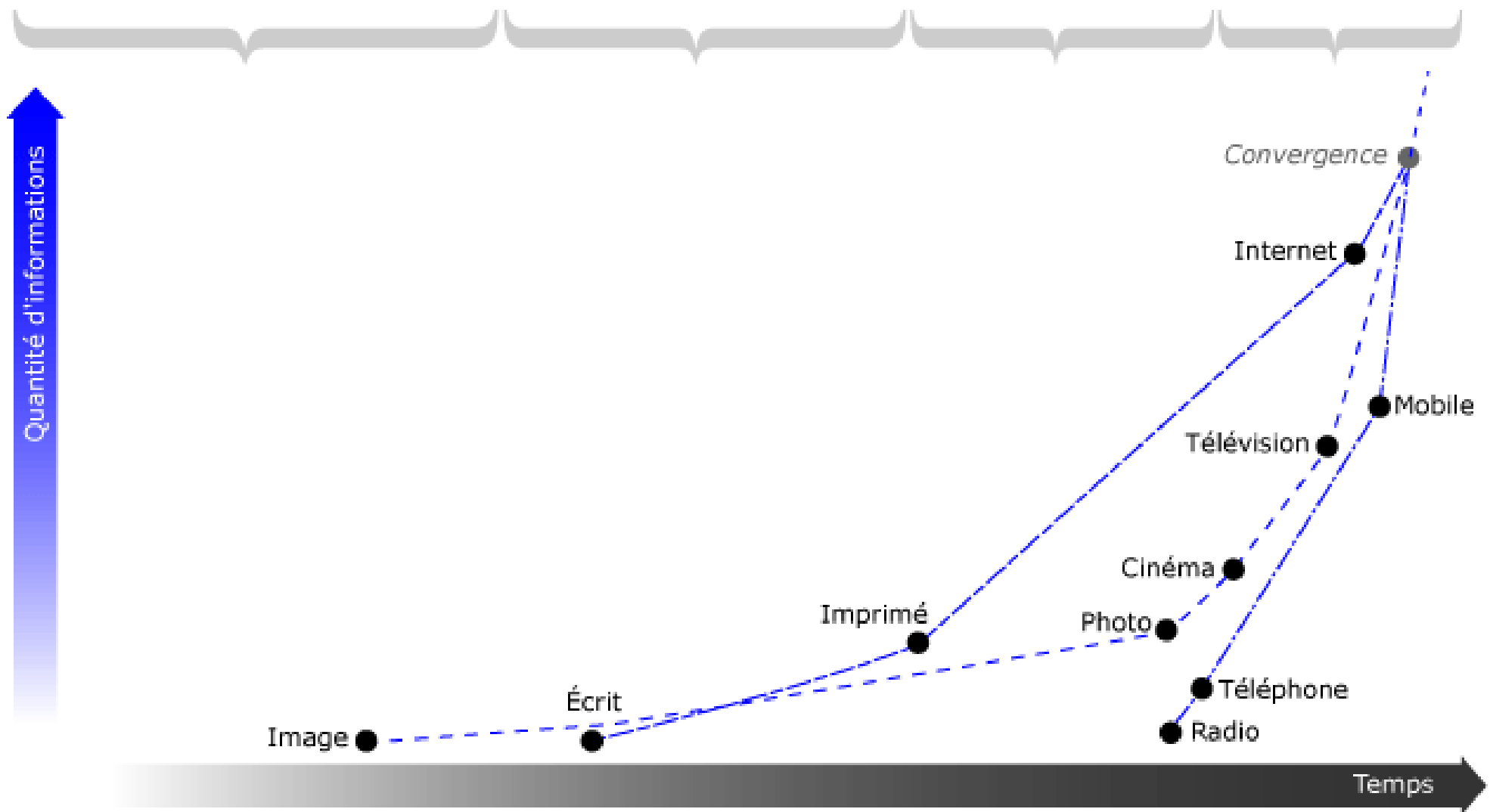
Région
Clan

Renaissance et Époque industrielle

Pays
Nation

Société du savoir

Système-monde
Groupes d'intérêts



Quelques « boulettes » »



- Tokyo stock exchange – déc. 2005
 - **Perte de 286 millions €**
 - Causes : un courtier a placé en bourse 610000 titres à 1 yen au lieu de 1 titre à 610000 yens lors de l'introduction en bourse d'une TPE – Jcom
-
- Trois touristes norvégiens attendus à Rhodes (Grèce) atterrissent à Rodez (France) suite à une erreur de saisie de leur commande internet
-
- Une réservation sur le site internet d'une compagnie low cost d'un vol retour décollant avant l'arrivée du vol aller

Coût de la non qualité



www.quares.fr

NASA

- 1999 : perte du satellite Mars Climate Orbiter
- Certains paramètres ont été calculés en paramètres anglo-saxons et transmis en l'état à l'équipe de navigation qui les attendaient en système métrique
- Coût pour le contribuable US : 125 millions de dollars

AIRBUS

- A la conception de l'A380, il a été prouvé que les difficultés concernant la phase d'industrialisation ont porté essentiellement sur le câblage électrique d'une partie du fuselage. La conception avait été menée avec des logiciels de versions différentes
- Coût pour EADS : des mois de retard de livraison et les pénalités inhérentes

Ex : Calcul du ROI d'une campagne marketing d'un opérateur téléphonique

- L'entreprise veut faire l'envoi d'une brochure annonçant un nouveau service à l'ensemble de ses clients

- Sa base Client contient des enregistrements dupliqués
- ratio de duplication 1.04

- Sa base Client contient des enregistrements multiples pour le même foyer
- ratio de foyer 1.11



Hypothèses		
Nombres de brochures envoyées	50000	
Coût total du programme	150 000 €	
Bénéfice moyen par vente	1 000 €	
Ratio de duplication	1	1.04
Taux de réponse	2 %	1.92 %
Ratio de foyer	1	1.11
Taux de conversion	20 %	18.02 %
Résultats		
Nombre de réponse	1000	962
Coût par réponse	150 €	156 €
Nombre d'acheteurs	200	173
Bénéfice total de la campagne	200 000 €	173 250 €
Retour sur Investissement	33.33 %	15.50 %

Et encore ...



- Un étudiant en médecine qui commence sa thèse en recueillant des données de t°- 37.1-37.2-37.5... et qui passe du jour au lendemain à la notion de – Fièvre / pas Fièvre
- Des données issues de « dossiers » qui n'étaient pas prévus pour ça au départ (ex : dossier judiciaire ou médicaux en SHS)
- Une carte maritime mal renseignée par le ministère des affaires étrangères
→ 15 mois de prison en Iran pour deux ressortissant Français et Allemand

Impact



www.quares.fr

Analyste

- Meilleure connaissance des problèmes et développement de nombreuses solutions

Producteurs
de données
et
praticiens

- Mise à disposition de méthodes pour mesurer et décrire la qualité des données

Utilisateurs
lambda

- Amélioration de la qualité (interne) de jeux de données

Impact

Enjeux pour la recherche

Scientifiques

- Le chercheur a besoin de juger la qualité de l'information et sa pertinence.
- Des réflexions visant le contrôle du processus de l'information sont engagées.

Economiques

- La qualité de l'information peut devenir un critère important de sélection pour bénéficier d'un financement national, européen ou international.

Sociétaux et environnementaux

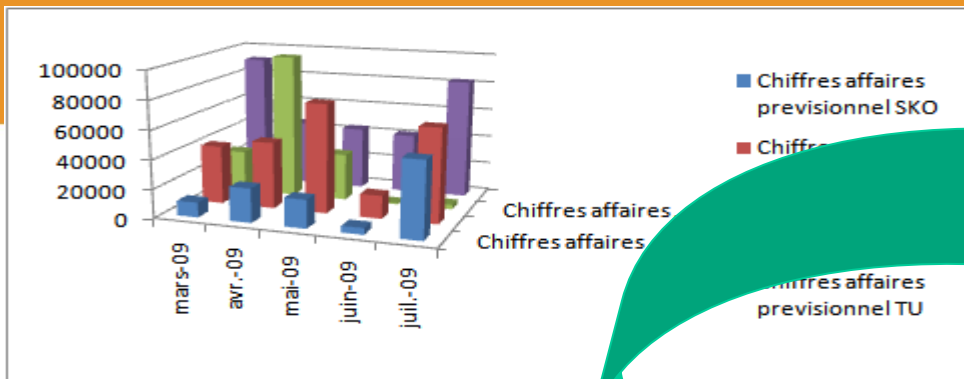
- Responsabilité du chercheur et des laboratoires vis à vis de leurs recherches

Quelques définitions

QuaRES, pour la Qualité en Recherche et en Enseignement Supérieur



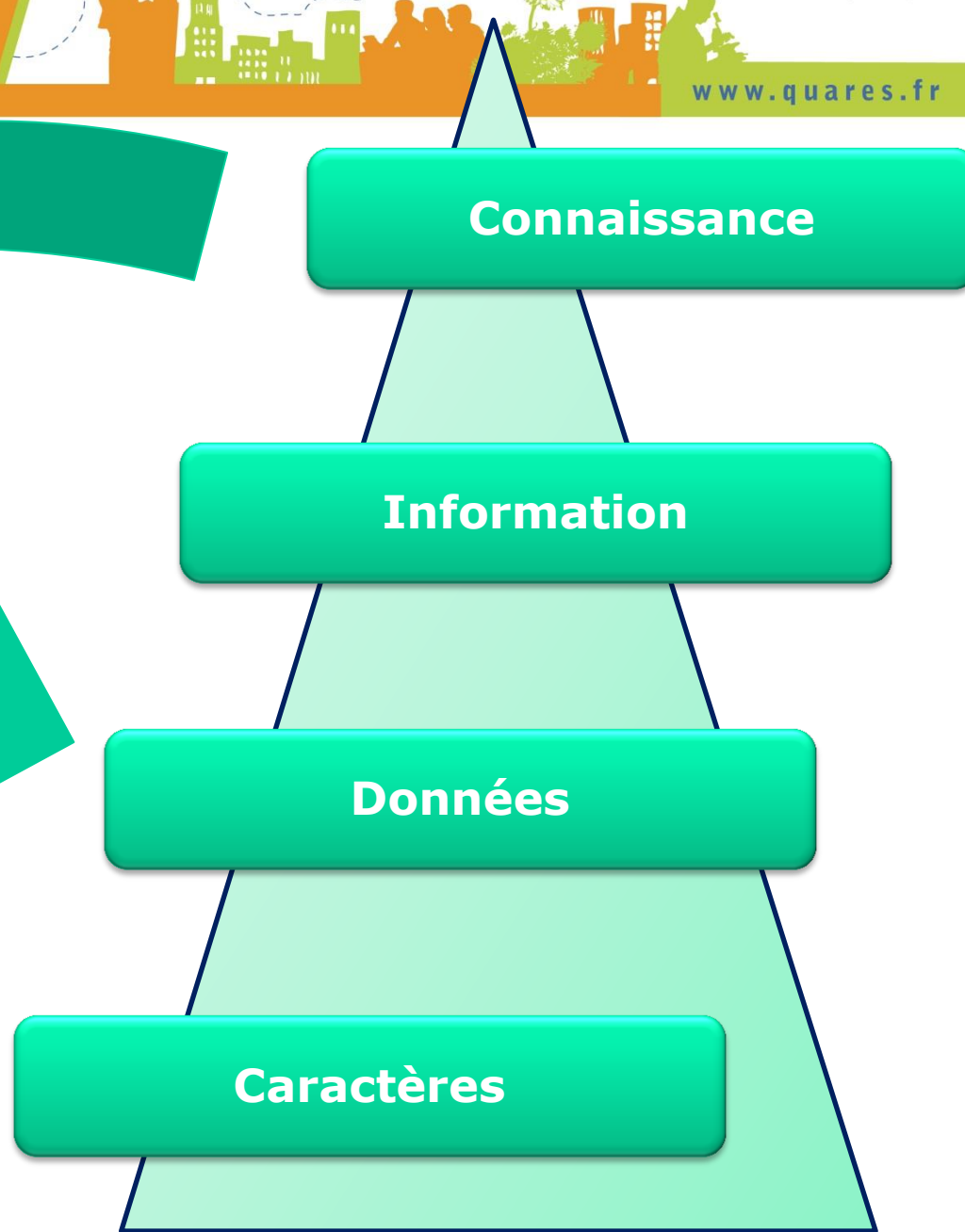
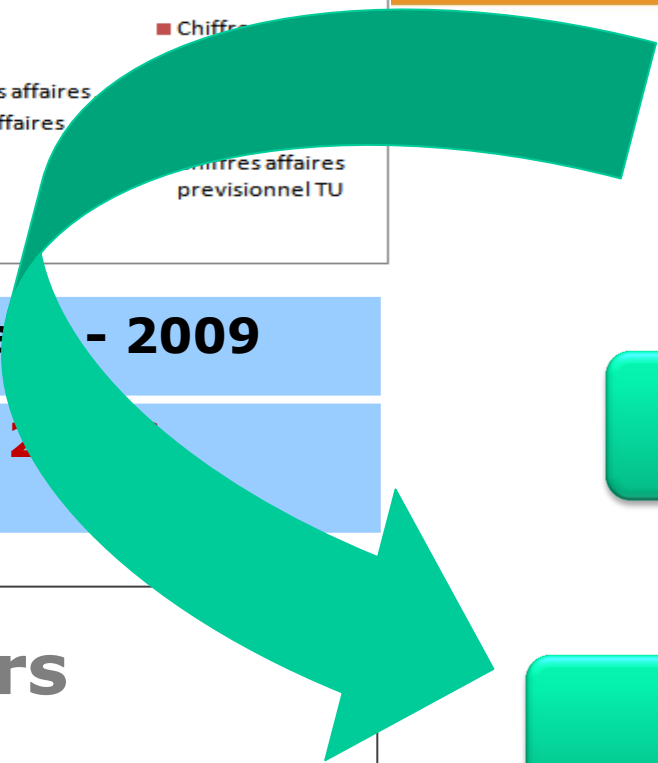
www.quares.fr



	Mars - 2009
Chiffre d'Affaires	13 245

13 245 Mars
2009

1 5 A 9
0 2 R €
3 0 4 M S





Terme **générique** définissant
à la fois

- les caractéristiques des **données complètes, fiables, pertinentes et à jour**
- mais aussi
- l'ensemble des **processus** qui permet de garantir ces caractéristiques

BRAINSTOMING

Ensemble réfléchissons à

10 causes de non qualité des données

Les 10 Principales causes de mauvaise qualité des données



www.quares.fr

- **1 - Erreurs typographiques et données non conformes**
- **2 - Offuscation d'informations**
- **3 - Informatique traîtresse et fichiers Excel**
- **4 - fusion de système d'information**
- **5 - Le changement est bon... Sauf pour la qualité de données**
- **6 - Code caché**
- **7 - Transition des transactions**
- **8 - Métamorphose de métadonnées**
- **9 - Définition de la qualité de données**
- **10 - Perte d'Expertise**

Par groupe

**Pour chaque causes de non qualité,
trouver ce que l'on peut mettre en œuvre
pour minimiser le risque de mauvaise
qualité des données**

1. Erreurs typographiques et données non conformes



- Malgré l'automatisation quasi systématique de l'architecture des données, des informations sont toujours saisies dans des formulaires Web et d'autres interfaces utilisées par les clients. **Les erreurs de saisie sont une source fréquente d'imprécision des données.** Les utilisateurs font parfois des **fautes de frappe.** Ils choisissent **la mauvaise entrée dans une liste.** Ils **saisissent la bonne donnée au mauvais endroit.**
- Etant donnée la complète liberté laissée pour renseigner un champ, les personnes saisissent des données de mémoire. Le nom du chercheur est-il Grainger, WW Granger, ou W. W. Grainger ? Idéalement, il devrait y avoir un ensemble de données de référence d'entreprise (métadonnées) afin que les formulaires permettent aux utilisateurs de trouver les bons vendeurs, noms de clients, villes, références, etc.

1. Erreurs typographiques et données non conformes



- Formation** – Assurer que les personnes saisissant les données connaissent leur impact sur les applications en aval.
- Définition des métadonnées** – En restreignant ce que les gens peuvent saisir dans un champ via une liste exhaustive, de nombreux problèmes peuvent être évités. Ces métadonnées (pour le nom des vendeurs, les références, etc.) peuvent devenir une partie de la qualité des données dans l'intégration de données, les applications métier et d'autres solutions.
- Monitoring** – Rendre publics le résultat des données mal saisies et féliciter les personnes ayant saisi des données correctement.
- Validation en temps réel** – En plus des formulaires, des outils de validation de qualité de données peuvent être implémentés afin de valider des adresses, des adresses e-mail ainsi que d'autres informations importantes lors de la saisie. Assurez-vous que votre solution de qualité de données puisse diffuser la qualité de données dans des environnements de serveurs d'application, dans le cloud ou dans un Enterprise Service Bus (ESB).

2. Offuscation d'informations



Les erreurs lors de la saisie de données ne sont pas toujours faites par erreur. Combien de fois les gens ne donnent-ils pas des informations incomplètes ou incorrectes afin de protéger leur vie privée ? Si cela n'a aucune incidence sur les personnes saisissant les données, elles ont tendance à les falsifier

Même si les personnes saisissant les données souhaitent le faire correctement, c'est parfois impossible. Si un champ n'est pas disponible, un autre champ est souvent utilisé. Cela peut causer des problèmes de qualité de données, tels que des numéros de TVA dans le champ du nom, ou des coordonnées dans le champ des commentaires.

2. Offuscation d'informations



- Récompense** – Offrir une prime aux gens saisissant des données personnelles correctement. Cela doit être centré sur les gens saisissant des données de l'extérieur, par exemple via les formulaires Web. Les employés ne doivent pas avoir besoin d'une récompense pour faire leur travail. Le type de récompense dépendra de l'importance d'avoir les informations correctes.
- Accessibilité** – En tant qu'expert en charge de l'arbitrage de données, soyez ouvert et acceptez les critiques des utilisateurs. Soyez à l'écoute lorsque les changements de processus nécessitent un changement de technologie. Si vous n'êtes pas accessible, les utilisateurs chercheront tous les moyens de valider leur formulaire, même incorrect.
- Validation en temps réel** – En plus des formulaires, des outils de validation de qualité de données peuvent être implémentés afin de valider des adresses, des adresses e-mail ainsi que d'autres informations importantes lors de la saisie.

3. Informatique traîtresse et fichiers Excel



Un renégat est une personne qui déserte et trahit un ensemble de principes d'une organisation. C'est exactement ce que font sans le savoir certains chefs d'entreprise impatients en déplaçant leurs données dans et hors de leurs solutions métier, bases de données, etc.

Plutôt que d'attendre de l'aide d'équipes informatiques professionnelles, des équipes métier zélées peuvent décider de créer leur propre ensemble d'applications locales, sans connaissance informatique particulière. Si l'application doit répondre aux besoins immédiats du département, il est peu probable qu'elle soit conforme aux standards de données, de modèles de données ou d'interfaces.

La base de données doit commencer par faire une copie d'une base de données approuvée vers une application locale sur le bureau des membres de l'équipe. D'importants morceaux de données stockés dans des feuilles de calcul Excel, dénommés « spreadmarts », sont facilement duplicables sur le bureau des membres de l'équipe. Dans ce scénario, vous perdez le contrôle des versions et des standards. Il n'y a pas de sauvegarde, de versionnement ou de règle métier.

3. Informatique traîtresse et fichiers Excel



- Culture d'entreprise** – Il devrait y avoir des conséquences pour les personnes propageant ces spreadmarts, les dissuadant de créer des applications de données locales.
- Communication** – Eduquez et formez vos employés sur les conséquences négatives des spreadmarts.
- Gestion de « Small data »** – Il est crucial d'avoir des outils permettant à des utilisateurs métier et aux professionnels informatique de gérer les données.
- Les solutions de **Master Data Management (MDM)** permettent de combler les lacunes entre les applications informatiques coûteuses et limitées et la gestion métier effective des données.
- Verrouiller les données** – Le but est d'atteindre une culture d'entreprise où créer des spreadmarts sans sanction n'est pas possible. Certaines entreprises ont trouvé le succès en verrouillant les données afin de les rendre plus difficiles à exporter.

4. Après la fusion



www.quares.fr

- Les fusions augmentent le risque d'erreurs de qualité de données car elles se déroulent en général rapidement et ne sont pas prévues par les départements informatiques.
- Presque immédiatement, il y a une certaine pression pour consolider et raccourcir le planning. La consolidation inclura probablement le besoin de partager des données dans un ensemble varié d'applications disjointes. De nombreux raccourcis sont pris pour rendre cela possible, ce qui comprend souvent des risques connus ou inconnus pour la qualité de données.
- Parmi les priorités du planning raccourci, la fusion des départements informatiques peut provoquer un conflit culturel et différentes versions de la vérité. De plus, des fusions peuvent créer une perte d'expertise lorsque des personnes clés partent au milieu du projet pour chercher de nouvelles aventures.

4. Après la fusion



- Conscience d'établissement** – Lorsque c'est possible, la direction doit ordonner la division du travail afin d'éviter des conflits de culture et la rétention de l'information par les gens assoiffés de pouvoir.
- Document** – Votre projet informatique doit perdurer, même si l'équipe entière part, se disperse ou est écrasée par un bus en traversant la rue. Vous pouvez réussir à le maintenir avec une bonne documentation de l'infrastructure.
- Consultants externes** – Le management doit savoir qu'il y a du travail supplémentaire à faire et que des conflits peuvent émerger après une fusion. Des consultants peuvent fournir la continuité nécessaire pour réussir la transition.
- Gestion de données agile** – Les solutions de data management permettent de garder une flexibilité, vous donnant ainsi la possibilité de diviser et conquérir la charge de travail sans licence coûteuse d'applications commerciales.

5. Le changement est bon... Sauf pour la qualité de données



Les entreprises subissent des changements de leurs processus métier afin de s'améliorer. C'est une bonne chose, n'est-ce pas ? Les principaux exemples comprennent :

- Expansion de l'entreprise sur de nouveaux marchés
- Nouveaux accords de partenariats
- Nouvelles lois de régulation sur le reporting
- Reporting financier à une entreprise mère
- Réduction d'effectifs

Si la qualité de données est définie comme l'adaptation aux besoins (« fitness for purpose »), que se passe-t-il lorsque l'objectif change ? C'est cette nouvelle utilisation de données qui apporte des changements dans le niveau perçu de qualité de données, même si les données sous-jacentes sont les mêmes. Il est naturel que les données changent. Lorsqu'elles le font, les règles de qualité de données, les règles métier et les couches d'intégration de données doivent également changer.

5. Le changement est bon... Sauf pour la qualité de données



- Gouvernance des données – En mettant en place une équipe de gouvernance de données transverse aux métiers, vous aurez toujours une équipe qui contrôlera les changements que subit votre entreprise et examinera leur impact sur les informations. Cela devrait être dans la charte d'une équipe de gouvernance de données.
- Communication – Une communication régulière et un modèle de métadonnées bien documenté simplifient le changement.
- Flexibilité des outils – L'un des défis lorsque vous achetez des outils de qualité de données embarqués dans des applications d'entreprise est qu'ils peuvent ne pas fonctionner dans toutes les applications d'entreprise. Lorsque vous choisissez des outils, soyez sûr qu'ils soient suffisamment flexibles pour fonctionner avec les données de n'importe quelle application et que l'entreprise soit attachée à la flexibilité et à l'ouverture.

6. Code caché



Les bases de données commencent rarement leur vie en étant vides. Le point de départ est généralement une conversion de données à partir d'une source de données déjà existante. Le problème étant que, si les données peuvent fonctionner parfaitement dans l'application source, elles peuvent échouer dans la cible. Il est difficile de voir tout le code applicatif maison et les processus spéciaux qui se déroulent derrière les données, sauf si vous les profilez.

6. Code caché



- Profilez tôt et souvent – Ne partez pas du principe que vos données sont adaptées à l’objectif visé parce qu’elles fonctionnent dans l’application source. Le profiling vous donnera une évaluation exacte de la forme et de la syntaxe des données dans la source. Il vous permettra également de savoir la quantité de travail que vous devrez fournir pour les faire fonctionner dans la cible.
- Appliquez des outils de qualité de données lorsque c’est possible – Plutôt que du code personnalisé dans l’application, une meilleure stratégie est de laisser les outils de qualité de données appliquer les standards. Les outils de qualité de données appliquent des standards d’entreprise de manière uniforme, ce qui simplifie le partage des données.

7. Transition des transactions



De plus en plus de données sont échangées entre les systèmes via des interfaces en temps réel (ou quasiment en temps réel). Dès que les données entrent dans une base de données, cela déclenche des procédures nécessaires à l'envoi de transactions à d'autres bases de données en aval. L'avantage est la propagation immédiate des données à toutes les bases de données correspondantes.

Mais que se passe-t-il lorsque les transactions tournent mal ? Un système qui fonctionne mal peut causer des problèmes aux applications en aval. En fait, même un petit changement dans un modèle de données peut poser des problèmes.

7. Transition des transactions



- **Vérifications des schémas** – Effectuez des vérifications de schémas dans les flux de vos Jobs afin de vous assurer que vos applications en temps réel produisent des données cohérentes. Les vérifications de schémas vont effectuer des tests simples afin de vérifier que vos données sont complètes et correctement formatées, avant de les charger.
- **Monitoring de données en temps réel** – Le niveau supérieur aux vérifications de schémas est le monitoring de données proactif avec des outils de profiling et de monitoring de données. Les outils comme de data management vous assurent que vos données contiennent le bon type d'informations. Par exemple, si vos références ont toujours une forme et une longueur particulières et qu'elles contiennent un ensemble fini de valeurs, toute variation de cet attribut peut être monitorée. Lorsque des variations se produisent, le logiciel de monitoring vous envoie une notification.

8. Métamorphose de métadonnées



Le référentiel des métadonnées doit pouvoir être partagé entre de nombreux projets, avec un audit régulier sur l'utilisation et l'accès au référentiel. Par exemple, votre entreprise peut avoir des références et des descriptions qui sont universelles pour le CRM, pour la facturation, pour les systèmes ERP, etc. Lorsqu'une référence devient obsolète dans le système ERP, le système CRM doit le savoir. Les métadonnées changent et doivent être partagées.

En théorie, documenter en totalité ce qu'il se passe dans la base de données et combien sont interdépendants les nombreux processus vous permettrait d'atténuer complètement le problème. Les descriptions et les références doivent être partagées entre toutes les applications concernées. Pour commencer, vous pouvez analyser les implications relatives à la qualité de données de tout changement dans le code, les processus, la structure des données, ou les procédures de collection des données, ce qui vous permet d'éliminer les erreurs de données inattendues. En pratique, cette tâche représente un énorme travail.

8. Métamorphose de métadonnées



- **Modèles de données prédéfinis** – De nombreuses entreprises possèdent des définitions très simples du contenu de chaque ensemble de données. Par exemple, l'industrie automobile suit certains standards ISO 8000. L'industrie de l'énergie suit les standards Petroleum Industry Data Exchange (PIDX). Cherchez un modèle de données dans votre domaine pour vous aider.
- **Gestion de données agile** – La réussite de la gouvernance de données se fait en commençant petit et en construisant un processus qui règle d'abord les problèmes les plus importants du point de vue métier. Vous pouvez tirer parti des solutions agiles afin de partager des métadonnées et de mettre en place des processus facultatifs à travers l'entreprise.

9. Définition de la qualité de données



De plus en plus d'entreprises reconnaissent le besoin de qualité de données, mais il y a différents moyens de nettoyer des données et d'améliorer leur qualité. Vous pouvez :

- Ecrire du code et nettoyer manuellement
- Manipuler la qualité de données dans l'application source
- Acheter des outils pour nettoyer les données

Considérez ce qu'il se passe lorsque vous avez au moins deux de ces types de processus de qualité de données ajustant les données. L'équipe commerciale a une définition du client, la comptabilité en a une autre. A cause des différents processus, les deux équipes ne s'accordent pas sur le fait que les deux enregistrements sont des doublons.

9. Définition de la qualité de données



- Outils de standardisation** – Choisissez des outils qui ne sont pas liés à une solution particulière. De la qualité de données uniquement dans SAP, par exemple, ne va pas améliorer vos ensembles de données Oracle, Salesforce et MySQL. Lorsque vous sélectionnez une solution, prenez-en une capable d'accéder à toutes les données, n'importe où et n'importe quand.
- Gouvernance de données** – En mettant en place une équipe de gouvernance de données transverse, vous avez les bonnes personnes au bon endroit afin de définir un modèle de données commun.

10. Perte d'Expertise



Dans presque tous les projets ayant de très grands volumes de données, il y a une personne dont l'expertise concernant les données héritées est exceptionnelle. Ce sont ces gens qui comprennent pourquoi la date d'embauche de certains employés est stockée dans le champ de la date de naissance et pourquoi certains attributs de noms contiennent également des numéros de TVA.

Les données peuvent être une sorte d'historique pour une entreprise. Elles peuvent provenir de systèmes hérités. Dans certains cas, la même valeur dans le même champ signifie quelque chose de totalement différent dans un autre enregistrement. Avoir conscience de ces anomalies permet aux experts d'utiliser les données correctement.

10. Perte d'Expertise



- Profiler et monitorer – Profiler les données vous permet d'identifier la plupart de ces types de problèmes. Par exemple, si un numéro de TVA se trouve dans un champ de nom, l'analyse vous permet de le remarquer rapidement. Le monitoring prévient les récurrences.
- Documenter – Même s'ils peuvent être réticents par peur de perdre la sécurité de leur emploi, assurez-vous que les experts documentent toutes les anomalies et les transformations devant advenir à chaque déplacement des données.
- Utiliser les consultants – Les employés experts peuvent être si précieux et si occupés qu'ils n'ont pas le temps de documenter les anomalies héritées. Les entreprises de consulting externes sont généralement très bonnes pour documenter les problèmes et fournir une continuité entre l'héritage et les nouveaux employés.

4 facteurs définissant l'information

Accessibilité

Disponibilité des systèmes

Disponibilité des transactions

Privilèges et droits d'accès

Utilité

pertinence

Duplication

Usage

Opportunité

- Actualisation des sources
- actualisation des cibles
- Volatilité

Crédibilité

Standardisation

Cohérence

Exactitude

Intégralité

Interprétabilité

Syntaxe

Sémantique

Contrôle des versions

Alias

Origine

6 dimensions définissant la qualité de l'information

Pertinence

Satisfaction du besoin client

Les données doivent éclairer les utilisateurs sur les questions les plus importantes à leurs yeux

Répondre au besoin compte tenu des contraintes en matière de ressources

Exactitude

Dans quelle mesure l'information décrit bien le phénomène qu'elle doit mesurer

Définie par rapport aux estimations statistiques erronées

Décomposée en composantes

- De biais : erreurs systématiques
- De variance : erreurs aléatoires
- D'erreurs principales susceptibles de mener à des données inexactes (Couverture, échantillonnage, non réponses, réponses)

Rapidité de diffusion

Délai entre la fin du recueil de l'information et la date à laquelle les données sont disponibles

Souvent en compromis avec l'exactitude des données

Accessibilité

Facilité avec laquelle on peut se procurer les données

Facilité avec laquelle on constate que l'information existe

Caractère approprié de la présentation de l'information

coût d'accès à l'information

Interprétation

Disponibilité de renseignements supplémentaires

Disponibilité de métadonnées nécessaires à l'exploitation et à l'interprétation

Classification et concepts sous-jacents

Méthodes de collectes de données

Indicateurs de l'exactitude des données

Cohérence

Liaison avec d'autres renseignements statistiques dans un cadre analytique au fil du temps

Utilisation de concepts de classification, de populations cibles

L'utilisation de méthodes et de techniques communes pour le recueil de l'information

Indicateurs et mesures

Critères

Opportunité

Intégralité
Complétude

Cohérence

Exactitude

Interprétabilité

Standardisation
conformité

Duplication

Caractéristiques

L'âge des données est-il conforme aux besoins métiers ?

Est-ce que toutes les données sont disponibles ?

Quelles sont les sources des informations contradictoires ?

Les valeurs représentent-elles la réalité ?

Les données sont-elles compréhensibles par les utilisateurs ?

Quelles sont les données saisies, stockées ou affichées dans un format non standard ?

Quelles sont les données répétées ?

Indicateurs

Date de la collecte des données

Date du dernier traitement

Contrôle de la version

Intégralité des valeurs optionnelles

Nombre de valeurs non renseignées

Nombre de valeurs par défaut par rapport à la moyenne

Vérification de la plausibilité

Valeur de la déviation standard

Fréquence des changements de valeurs

Réaction (feedback) des clients

Valorisation des données utilisateur

Violation de domaines

Certificat de conformité

Nombre d'enregistrements dupliqués

Qui est concerné ?

Dirigeants

Premiers promoteur de la qualité de l'information

Définition de la stratégie

Prises de décision

Pilotage

Responsables opérationnels

Garant de la qualité de l'information à sa création

Gestion au quotidien d'important volume de données

Interface entre la direction et le métier

Besoins d'éléments pertinents et synthétiques

Collaborateurs Opérationnels

Créateur et utilisateur de données correctes

Besoin d'information fiable

Besoin d'un accès rapide à la donnée

Responsable du Système d'Information

Garant de l'intégrité et de la disponibilité de l'information

Former

Conseiller

Aider

Accompagner

Gouvernance

Direction générale et directions opérationnelles

Comme dans toute démarche qualité la sponsorship par la direction de l'entité est impérative

Elles s'assurent que la démarche est lancée

Elles créent un comité qualité des données

Comité qualité des données

Responsable de la qualité des données de l'entité

Il s'assure que l'ensemble des projets intègre la gestion de la qualité des données dans leurs processus d'acquisition de transformation et de restitution des données

Il assure le suivi de la démarche et fait le point sur les différents axes d'améliorations

- Des enjeux majeurs pour la recherche
- La qualité de l'information :
Tout le monde est concerné !
- Une organisation à mettre en marche



« La qualité des données, en informatique se réfère à la conformité des données aux usages prévus, dans les modes opératoires, les processus, les prises de décision, et la planification »

(Joseph Juran)



Les bonnes décisions sont prises avec des informations pertinentes

Les données contribuent au succès de l'activité de l'entreprise. Leur qualité représente donc un enjeu critique pour l'entreprise dans les 3 étapes de leur cycle de vie :

- **Lors de la saisie**
- **Au cours des transformations et agrégations**
- **Pendant l'analyse et la présentation des résultats**

Questions à se poser :

- **Comment identifier les données de mauvaise qualité ?**
- **Quel impact a la non qualité sur l'organisation ?**
- **Comment mesurer la non qualité ?**
- **Comment corriger ces problèmes ?**

Rappel sur les données



SOURCES d'information

- **Source primaire**
 - **Publication à l'origine d'une information (article, brevet...)**
- **Source secondaire**
 - **Permet la diffusion d'une information, analyse cette information, compilation et « amélioration » par la mise en perspective**

TYPES d'information

- **Information informelle / formelle**
- **Information stratégique / tactique / opérationnelle**
- **Information blanche / grise / noire**

Rappel sur les données



- Sources **formelles** : transcrites sur un support (papier ou électronique)
- Sources **informelles** : sources « orales » qui viennent souvent du terrain
ex : force de vente, salons, manifestations, clients, fournisseurs, réseau relationnel,
« rapports d'étonnement »...
- Sources **ouvertes** ou **fermées**

Rappel sur les données



- Information **formelle**
 - Information écrite
 - Informations qui ont un support : papier numérique
 - Journaux spécialisés, revues professionnelles, magazines, cours, articles scientifiques, brevets, bases de données scientifiques et documentaires, Internet
 - Plutôt veille
- Information **informelle**
 - Information sans support
 - « Ce qui est dit »
 - Partenaires de l'entreprise, Réseaux d'experts internes/externes, Reverse-engineering, Foires et salons
 - Plutôt IE

Rappel sur les données...



- Information **opérationnelle**
 - Information très ciblée, précise, de faible volume
 - Information brute ou peu retraitée
 - A destination des techniciens, ingénieurs, opérateurs
 - Veille brevet et technique par exemple
- Information **tactique**
 - Volume d'informations plus élevé
 - Subissant un lourd traitement
 - Veille concurrentielle, tarifaire
- Information **stratégique**
 - Ne concerne pas directement l'entreprise
 - De grande ampleur
 - Permet d'établir des indicateurs
 - Veille marché, prospective, statistiques

Rappel sur les données...



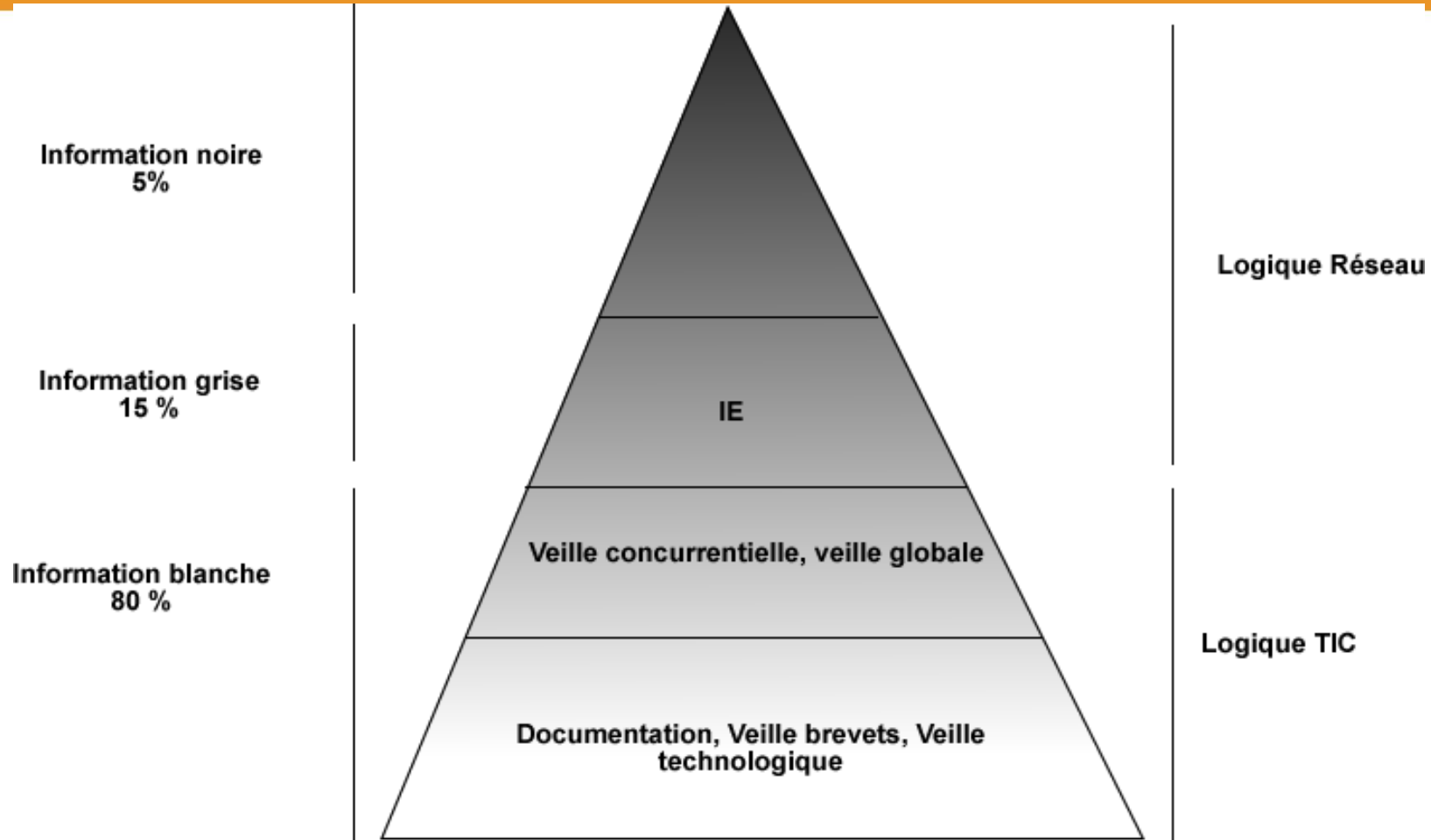
- Information **blanche**
 - Facilement accessible par tous
 - Peu de valeur
 - Nécessite tri et traitement important
 - Veille, bibliométrie, data-mining
- Information **grise**
 - Information difficilement accessible
 - A forte valeur
 - Souvent informelle
 - Indiscrétions, salons, ...
 - Intelligence économique
- Information **noire**
 - Information ne pouvant être acquise que de façon illégale
 - Information décisive pour l'entreprise
 - Espionnage industriel

Rappel sur les données



- L'information **blanche** :
Publique et accessible, ne fait l'objet d'aucune sécurisation particulière. Recherche "classique" dans les outils grands publics. Ce que chacun peut voir et recueillir notamment *sur le Web*.
- L'information **grise** :
Ne fait pas l'objet de publicité, mais on peut la trouver de manière indirecte ou détournée - Information *sensible d'accès légal*. Résulte d'un travail intellectuel. Techniques avancées de recherche et de traitement de l'information, groupe de discussion, liste de diffusion
- L'information **noire** :
Fait l'objet d'une *haute sécurisation* – Protégée par la loi ou le contrat. Réservée aux personnes autorisées. Relève de *l'espionnage industriel*

Rappel sur les données



La donnée : Rappel