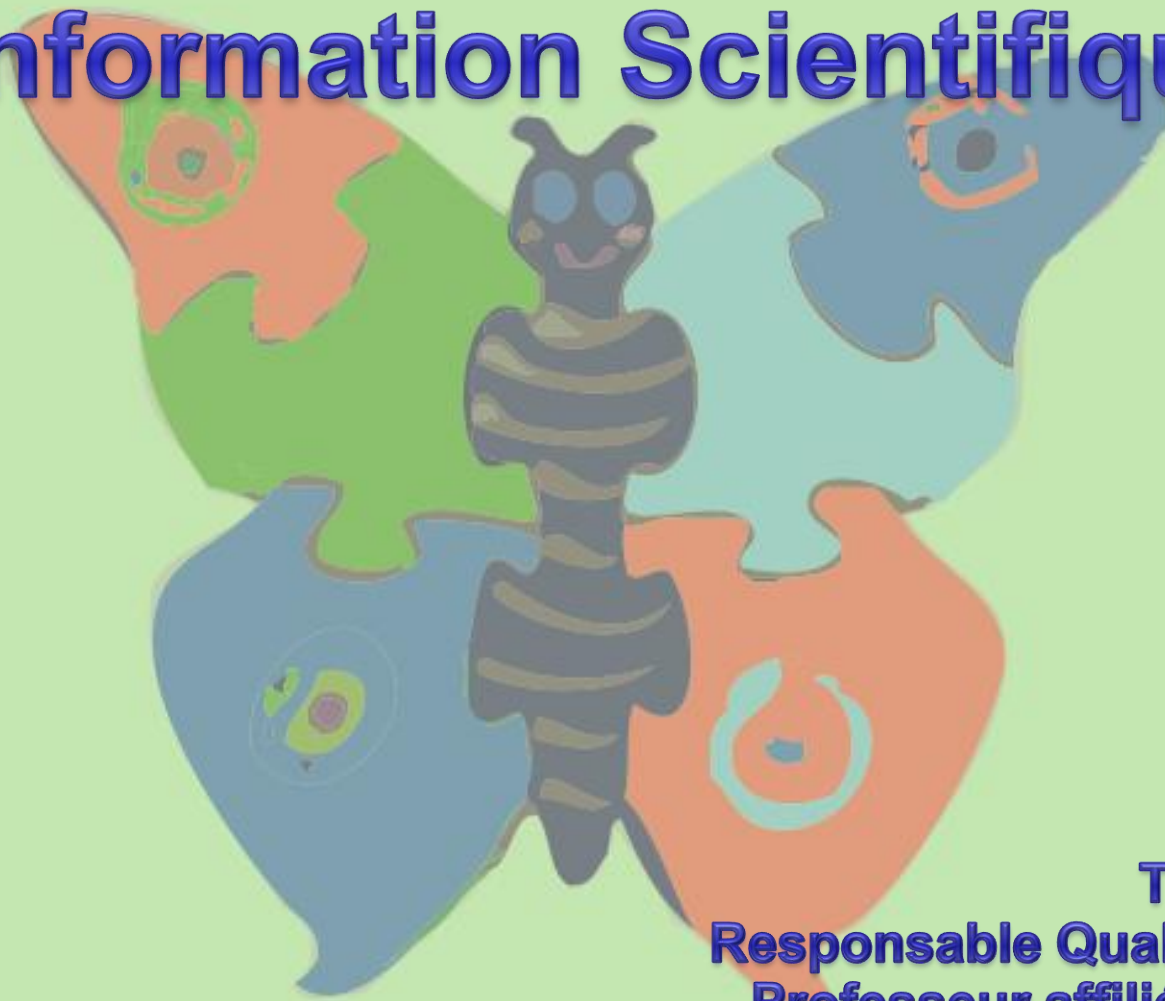


Qualité de l'Information Scientifique



Thierry BONTEMS
Responsable Qualité - UMR PACTE
Professeur affilié – ESC Grenoble

Un constat

- Plus d'informations ont été générées durant les 30 dernières années que pendant les 500 qui les ont précédées
- Les nouvelles technologies de type RFID génèrent un volume de données 10 à 100 fois supérieur à celui des technologies de type code-barres
- Aux Etats-Unis une étude⁽¹⁾ a estimé le coût de la mauvaise qualité des données à plus de 600 milliards de dollars par an

(1) « data quality and the bottom line » -The Data Warehousing Institute 2002



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

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

Audition de M. Louis Gallois- coprésident exécutif d'EADS du 22 novembre 2006



Introduction

Coût de la non qualité

Enjeux et Impact

Définitions et critères

Qui est concerné ?

Gouvernance

Conclusion

3

Coût direct de la non qualité des données

Ex : Calcul du ROI d'une campagne marketing d'un operateur 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

→ ROI qui varie du simple au double

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 %



Introduction

Coût de la non qualité

Enjeux et Impact

Définitions et critères

Qui est concerné ?

Gouvernance

Conclusion

4

Et chez nous ...

- 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

GIGO : GARBAGE IN → GARBAGE OUT



Introduction

Cout de la non qualité

Enjeux et Impact

Définitions et critères

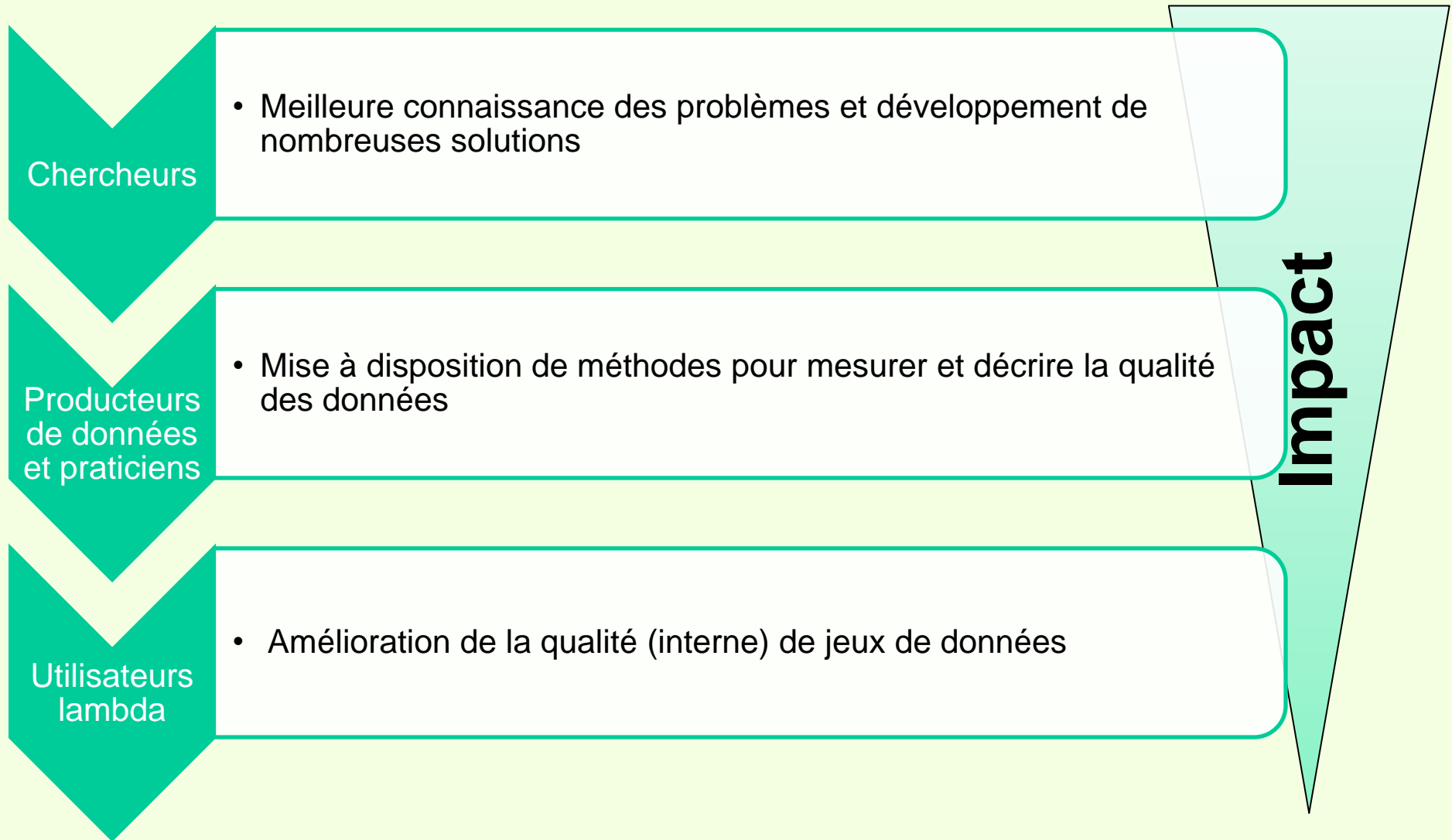
Qui est concerné ?

Gouvernance

Conclusion

5

Impact pour la recherche



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



Introduction

Coût de la non qualité

Enjeux et Impact

Définitions et critères

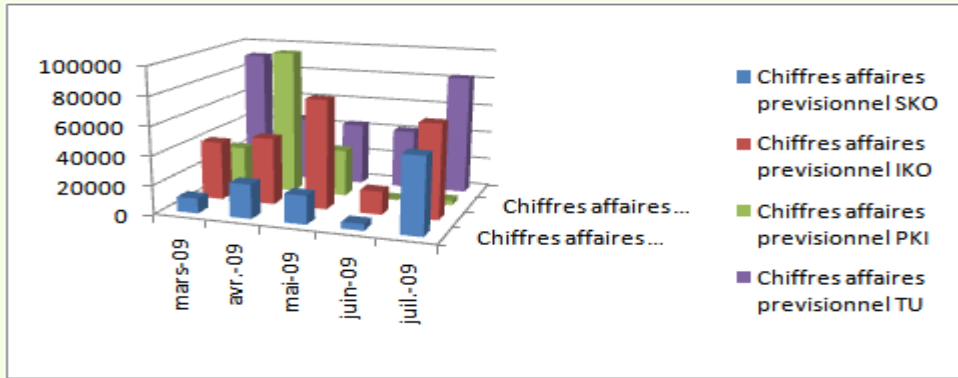
Qui est concerné ?

Gouvernance

Conclusion

7

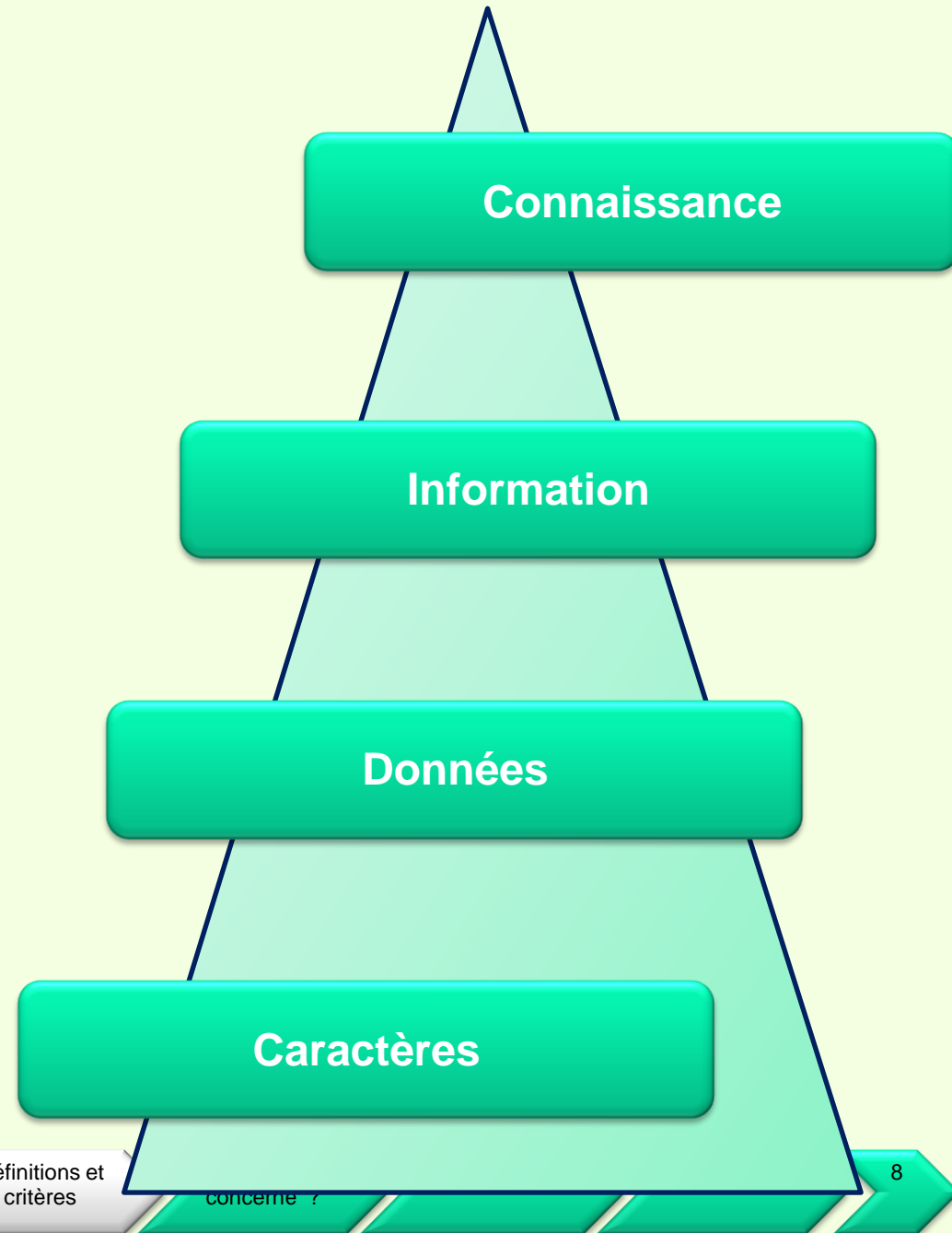
Quelques définitions



	Mars - 2009
Chiffre d'Affaires	13 245 €

13 245 Mars 2009

1 5 A 9
0 2 R €
3 4 M S



Qualité des données

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



Introduction

Coût de la
non qualité

Enjeux et
Impact

Définitions et
critères

Qui est
concerné ?

Gouvernance

Conclusion

9

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 au besoins métiers ?	Est-ce que toutes les données sont disponibles ?	Quelles sont les données 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

Conclusion

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



Ce dont on aurait pû parler !

- Intelligence Economique

- Business Intelligence

- Veille

- Analyse de risque

- Sécurité de l'information – ISO 27001

- ...



Merci de
votre attention

