



## Est-ce la qualité des données qui fait débat ? ou la compréhension des données et la définition des besoins ?

Henri Pornon

5 mai 2010

IETI Consultants – 17 boulevard des Etats-Unis - F-71000 Mâcon

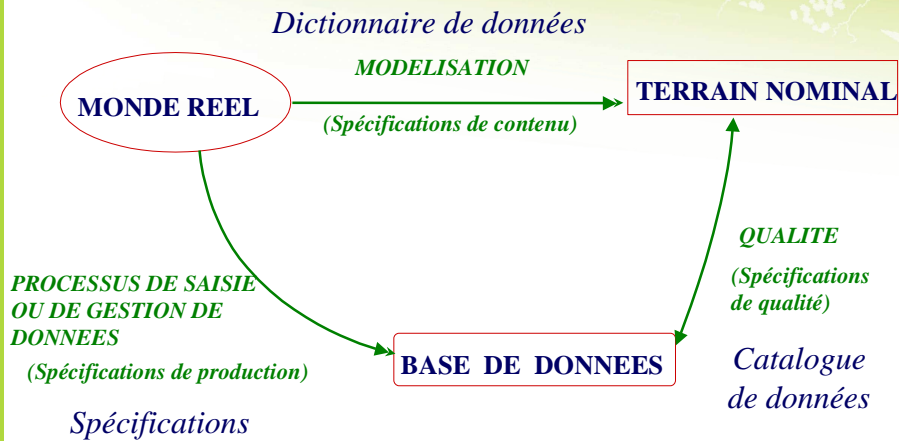
Tel : (0)3 85 21 91 91 - fax : (0)3 85 21 91 92

Email : [henri.pornon@ieti.fr](mailto:henri.pornon@ieti.fr) - Web : [www.ieti.fr](http://www.ieti.fr)

## ■ Avant-propos

- Quelques propos entendus ici ou là...
  - "Nous devons fournir des données de qualité irréprochable"
  - "Ces données sont de mauvaise qualité"
  - "Nos métadonnées contiennent toutes les informations utiles sur la qualité des données"
- Quelques constats sur la qualité des données
  - Des normes : ISO 19113 et 19114
  - Des outils : logiciels, méthodes, techniques, outils terrains..
  - Des documentations (ouvrage HERMES, CD ROM CERTU)
  - Des compétences
  - *On peut considérer que le sujet est maîtrisé du point de vue technique et conceptuel*
  - *Mais la qualité des données reste un problème*
- Pourquoi ?

## Rappel : concepts de la qualité



- ❑ Adéquation d'un produit ou service aux besoins d'un utilisateur

## Quelques définitions

- ❑ **Qualité** : ensemble des propriétés et caractéristiques d'un produit ou service, qui lui confèrent l'aptitude à satisfaire des besoins explicites ou implicites
- ❑ **Terrain nominal** : modélisation d'un domaine à l'aide de **spécifications (de contenu)** :
  - Vue du monde réel à travers les spécifications utilisées pour la saisie du lot de données géographiques
  - « Ce qui aurait dû être saisi »
  - un même domaine sera traduit en des terrains nominaux différents pour différents utilisateurs.
  - On parle aussi d'univers du discours.
- ❑ **Spécifications de production** : règles applicables au processus de production de la base de données
- ❑ **Spécifications de qualité** : niveau d'erreur acceptable dans la base de données par rapport au terrain nominal (lui-même décrit par les spécifications de contenu)

## ■ Qualité des données

### 4 situations problématiques

1. **Commande de données spécifiques.** Le client élabore un cahier des charges, contractualise avec un prestataire, contrôle la qualité des données livrées et constate qu'elles ne sont pas conformes.
2. **Commande de données sur étagères :** le client passe commande d'un produit sur étagère pour répondre à un besoin, contrôle la qualité des données livrées et constate qu'elles sont mal adaptée au contexte d'utilisation prévu
3. **Récupération de données sur une plate-forme partenariale :** les données récupérées auprès d'un partenaire sont inadéquates à l'usage projeté par l'utilisateur
4. **Intégration de plusieurs sources de données dans un observatoire du territoire ou un processus décisionnel.** Les sources sont hétérogènes car il n'y a pas correspondances entre les diverses définitions et les indicateurs produits à l'aide de ces sources ne sont pas considérés comme fiables

## ■ Commande de données spécifiques

### d'où peut venir le problème ?

- Spécifications techniques du client inadéquates ou insuffisamment documentées, maîtrise insuffisante du contexte par le client
- Incompréhension entre client et fournisseur sur la part implicite et la part explicite du besoin
  - "Numériser le POS / PLU" (sous-entendu : "en cohérence topologique avec la couche parcelle")
- Conflit entre client et fournisseur sur l'obligation de moyen ou de résultat et sur le devoir de conseil du fournisseur
- Prestation du fournisseur inadéquate, maîtrise des processus par le fournisseur

## ■ Commande de données standardisées d'où peut venir le problème ?

- ❑ Spécifications techniques du client inadéquates ou insuffisamment documentées, maîtrise insuffisante du contexte par le client
- ❑ Décalage entre besoin du client et spécifications de production du fournisseur
- ❑ Spécifications du produit
  - insuffisamment documentées (métadonnées insuffisantes)
  - ou mal comprises par le client (métadonnées peu compréhensibles, maîtrise insuffisante du contexte par le client)
- ❑ Produit non conforme aux spécifications du fournisseur

## ■ Récupération sur plate-forme d'où peut venir le problème ?

- ❑ Décalage entre besoin du récepteur et spécifications du producteur
- ❑ Métadonnées incomplètes ou insuffisantes ou peu compréhensibles par le client
- ❑ Maîtrise insuffisante du contexte par le client
- ❑ Données non conformes aux métadonnées

## ■ Intégration de plusieurs sources d'où peut venir le problème ?

- ❑ Décalage entre besoin du récepteur et spécifications du producteur
- ❑ Métadonnées incomplètes ou insuffisantes ou peu compréhensibles par le client
- ❑ Maîtrise insuffisante du contexte par le client
- ❑ Données non conformes aux métadonnées
- ❑ Données hétérogènes entre les divers producteurs

## ■ Problèmes de qualité interne

- ❑ Définition des besoins
  - Spécifications techniques inadéquates ou insuffisamment documentées (parfois, absence de cahier des charges)
  - Incompréhension entre fournisseur et client sur la part implicite / explicite du besoin
- ❑ Qualité des données
  - Prestation du fournisseur inadéquate, maîtrise insuffisante des processus
  - Produit non conforme aux spécifications de l'émetteur ou aux métadonnées
  - Données de divers émetteurs hétérogènes
  - Incapacité du client à contrôler le résultat de la prestation

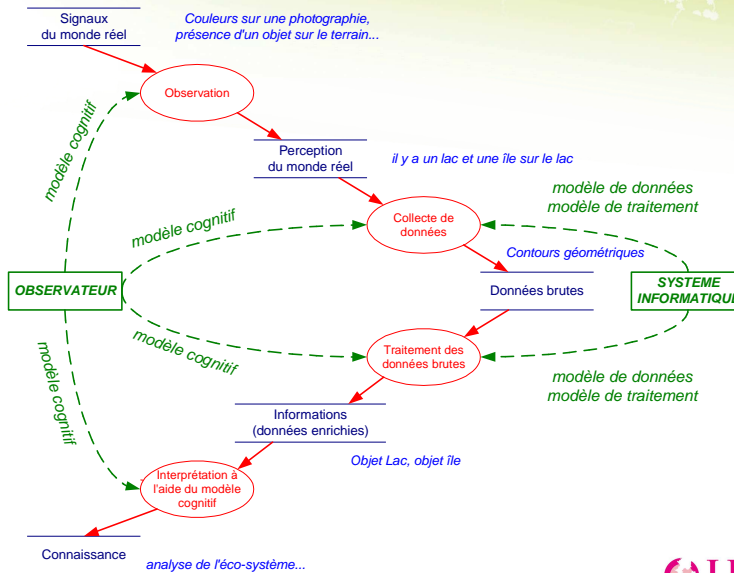
## ■ Problèmes de qualité externe

- ❑ Compréhension des données
  - Maîtrise insuffisante du contexte technique par le client
  - Spécifications du produit
    - insuffisamment documentées (métadonnées insuffisantes)
    - ou mal comprises par le client (métadonnées peu compréhensibles, maîtrise insuffisante du contexte par le client)
  - Problèmes de définitions liés à la sémantique des données
- ❑ Autres aspects
  - Décalage entre besoins du récepteur et spécifications de production de l'émetteur
  - Devoir de conseil du fournisseur, obligation de moyen ou de résultat

## ■ Rappel : une donnée n'est pas donnée mais construite

- ❑ Aucune donnée informatisée ne peut avoir le caractère d'universalité qu'on lui prête
- ❑ Toute donnée est produite :
  - Dans un contexte donné
  - Pour un usage donné
  - Par un ou plusieurs acteurs
  - En référence à un modèle cognitif implicite ou explicite, à une méthode et à des concepts
  - Dans un schéma conceptuel qui est une simplification du monde réel
- ❑ Le fait de standardiser des données ne les rend pas universelles, mais seulement homogènes et en général, les appauvrit du point de vue sémantique

## Du territoire à la connaissance



## Des solutions ?

### ☐ Métadonnées ?

- Pas besoin d'ouvrir le pot de confiture pour savoir ce qu'il y a dedans, mais qui sait lire les étiquettes et comprend les mentions ?

### ☐ Standardisation des données ?

- Facilite la compréhension et le partage des données ? Appauvrissement systémique ?

### ☐ Certification ?

- Améliore la qualité interne, mais pas externe
- Possible sur des données standardisées, produites dans des processus industriels ?

## ■ Pour améliorer...

- ❑ La qualité interne des données
  - Améliorer les cahiers des charges
  - Développer les compétences des maîtres d'ouvrage ou faire appel à des AMO
  - Réaliser de véritables contrôles qualité
- ❑ La qualité externe des données
  - Impossible ! Cela n'a pas de sens
  - Améliorer la compréhension des concepts, des données, des métadonnées
  - Trouver d'autres moyens d'information que les métadonnées plus accessibles aux non spécialistes ?



Merci de votre attention  
Des questions ?

IETI Consultants -17 boulevard des Etats-Unis - F-71000 Mâcon

Tel : (0)3 85 21 91 91 - fax : (0)3 85 21 91 92

Email : [henri.pornon@ieti.fr](mailto:henri.pornon@ieti.fr) - Web : [www.ieti.fr](http://www.ieti.fr)