

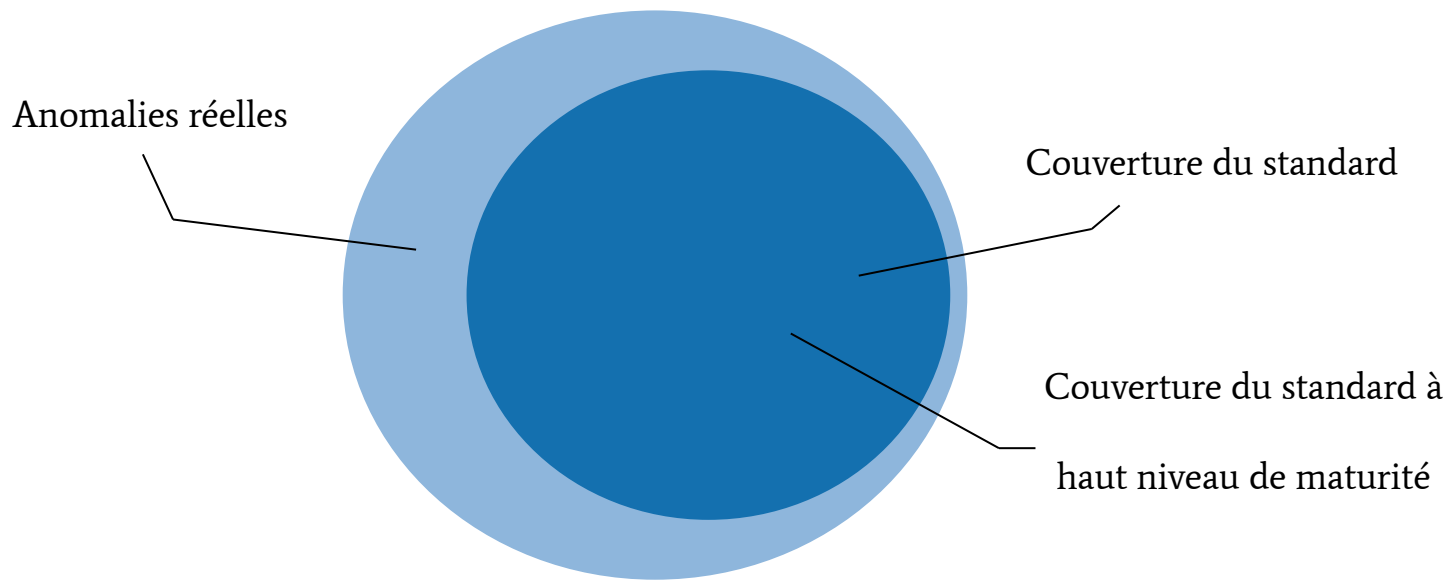
Standard de qualité de données et monitoring

Définition d'un standard de qualité

- Ensemble de règles de qualité de données, définis avec les producteurs/consommateurs de données
- 1 standard = 1 jeux de données = 1 ensemble d'usages similaires
- Utiliser nos connaissances sur la réalité pour vérifier la qualité des données



Le standard ne sera pas parfait

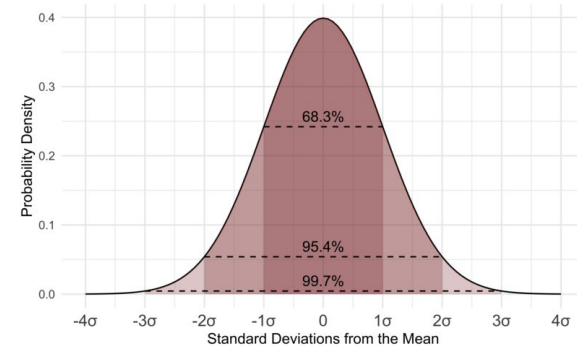


Indicateurs : probabilistes ou déterministes

Déterministes



Probabilistes



Dimensions de la qualité de données

- Pas de consensus sur le nombre ou la définition
- Utile pour agréger les indicateurs mais peut créer de la confusion si utilisées pour la définition

Exemple de standard

[Notebook]

Définitions des règles et indicateurs

1. Profilage des données et analyse de la documentation
2. Interview/workshop avec producteurs et consommateurs
3. Accord sur le standard

Situation dégradée : Standard générique

Générique

Facile à déployer



Pertinence



Spécifique

Facile à déployer



Pertinence



Indice de qualité de données

- Indice par dimension
- Indice global
- Pondérée vs poids égaux
- Attention au changements de définitions !



Exemple d'indice

[Notebook]

Automatisation

- Va dépendre de la stack technique
- Dans la plupart des cas, la difficulté technique est faible

A la prochaine !

**Si la vidéo t'as plu, laisse un like et
commentaire**

A la prochaine !

Si la vidéo t'as plu, laisse un like et
commentaire