

Qualité des données dans l'IoT

Exemple de la Ville Intelligente

Le contexte : la ville intelligente (SmartCity)

- La smartcity est un concept de plus en plus répandu
- Profusion de données disponibles en Opendata
- Les données sont au centre du monde de l'IoT
 - ▷ C'est une source d'extraction d'informations
 - ▷ C'est un moyen de communication.

⇒ Quid de la **Qualité de ces données** ?

Définition de la Qualité

La qualité des données

La Qualité des données fait référence à la façon dont les données répondent aux exigences des consommateurs de données

(Wang et Strong, 1996; Batini et Scannapieco, 2006)

- La qualité des données est centrée sur l'utilisateur
 - En opposition à une vision intrinsèque et une dimension d'exactitude
 - Chaque utilisateur définit ses propres critères
- ⇒ Ces critères sont connus sous le terme de **Dimensions** de la qualité des données

Dimensions de la Qualité - 4 catégories (Wang et Strong, 1996)

- **Intrinsèque** : qui est innée ou qui existe de manière inhérente dans les données.
- **Contextuel** : par rapport au contexte des tâches utilisant les données. Délais, exhaustivité, volume des données
- **Représentation** : dans quelle mesure les formats de données sont représentatifs et compréhensibles.
- **Accessibilité** : dans quelle mesure les données sont accessibles (et en même temps sécurisées) pour les consommateurs de données. Accessibilité, sécurité de l'accès

Facteurs qui affectent la Qualité des données pour l'Internet des objets

- **Échelle de déploiement** : Le nombre considérable de dispositifs accumule les risques d'erreur.
- **Contraintes liées aux ressources** : énergie, capacités de calcul
- **Réseau** : pertes de connexion et de paquets. Les objets ne sont capables de transmettre que des messages de petite taille en raison de leurs ressources limitées.
- **Capteurs** : faible précision, incohérences
- **Environnement** : endroits extrêmes, neige, poussières, maintenance difficile
- **Vandalisme** : le déploiement dans la nature ouverte en fait une proie facile pour le vandalisme des humains et des animaux
- ...

Amélioration de la Qualité

- **Détection de données aberrantes** : considérées comme des événements dont la probabilité d'occurrence est extrêmement faible.
- **Interpolation** : L'interpolation consiste à déduire les valeurs manquantes en se basant sur d'autres valeurs (disponibles).
- **Intégration des données** : sources hétérogènes, l'intégration basée sur la sémantique (norme ISO 8000)
- **Déduplication** : mécanisme de compression
- **Nettoyage** : Terme qui peut englober les items précédents

Application à la ville Intelligente

- De nombreuses données sont accessibles en Opendata ⇒ mise en forme à façon
- Possibilité d'intermédiaires de fournisseurs de données (Data Brokers)
- Les données sont souvent transformées ⇒ nécessité de transparence sur les techniques et/ou algos utilisées
- Nécessité de connaître la provenance des données ⇒ biais possible
- C'est la **Qualité des données** qui doit intégrer la **Gouvernance des données** ou la Gouvernance qui doit intégrer la Qualité ?

Qualité des données dans l'IoT

Exemple de la Ville Intelligente, 1^{er} Juin 2022

Vincent Lalanne - Manuel Munier

Merci pour votre attention.

`vincent.lalanne@univ-pau.fr`

`manuel.munier@univ-pau.fr`



<https://lalannev.perso.univ-pau.fr/>

<https://munier.perso.univ-pau.fr/>