

Acquisition, agrégation, partage

enjeux éthiques du Big Data

Les données

Une fois acquises/agrégées, on ne revient jamais en arrière (Pandora's box et persistance)

L'agrégation de données peuvent changer le niveau de sécurité qui y est attaché. (adresse/étage ≠ adresse/étage/code d'accès)

Quelle que soit la manière dont elles sont traitées, les données collectées doivent être comprise par le client (priorité)

La création de jeux de données (dataset) est essentielle, pour leurs conservations, partages et agrégations potentiels.

Le partage des données assure l'innovation ↔ une licence virale garantit le partage des nouveautés trouvées (OpenData)

Exemple1 : Air Vera City (secteur privé)

Capteurs statiques et mobiles

Agrégation avec des données géographiques

Observation du niveau de pollution

OpenSource, OpenHardware, OpenData (à la demande)

Exemple 2 : OpenData.Swiss (secteur institutionnel)

Projet de l'Office fédéral de la statistique

Départements fédéraux, cantons et villes

Des jeux de données (datasets) variés, librement accessibles dans de nombreux formats

Permet l'innovation dans de nombreux projets, civils, institutionnels et privés

Données en temps réel, données annuelles, données patrimoniales

Exemple 3 : EchoTopos (société civil)

Données soumises individuellement

Agrégation de données publiques, de données civiles et de données personnellement soumises par l'utilisateur

Partage par application

Étapes de travail

Acquisition : Quelles données sont déjà collectées ? Sont-elles disponibles ? Quels outils pour les collectés ? Fréquence de la collecte/stockage des données ? Une politique de sécurité est-elle en place (transmission et stockage) ?

Observation : rechercher les points dans les données qui peuvent être dignes d'intérêt, c'est le travail du data scientist/analyst → puis étiqueter les premiers résultats → définition des niveaux de sensibilités des données

Priorités : Mettre la priorité sur les données/informations qui correspondent au résultat initial attendu → préparer les premiers outils de visualisation des données → gestion des accès

Assigner : Définir les notifications → création des jeux de données (dataset) → créer les visualisations nécessaires → établissement des protocoles de partages

Petites itérations = tâches bien définies = résultats

Conclusion

L'éthique est le résultat des circonstances que l'on a et des limites que l'on se fixe. En big data, il est difficile de prévoir les conséquences de l'usage des données. Une bonne compréhension des données collectées et des analyses effectuées permet d'appréhender ces conséquences.

Le libre accès aux données garantit une innovation homogène et permanente en évitant la stagnation du monopole. Le partage continu des résultats évite les nombreuses répétitions associées à la R&D. (Pas de ré-invention de la roue).

Une fois cette compréhension acquise la présentation et la sécurisation des données se fait naturellement. Les politiques de sécurisation, partages et acquisitions peuvent être ainsi fixés, créant un cadre éthique.