
Qualité des données ouvertes

Exercice 1.1 - Mise en œuvre des Données Ouvertes

L'initiative données ouvertes ou Open Data permet de rendre disponible un certain nombre de données publiques ou privées, couvrant un large spectre de domaines.

Il existe aujourd'hui de nombreuses sources disponibles, en France et dans le monde, dans des formats standards (csv, xml, json, etc.) ouvrant la voie à une large gamme d'applications (voir les références).

L'objectif de ce travail est de réaliser une application de type médiateur intégrant un sous-ensemble des ces sources **choisies judicieusement**.

Travail à faire

- Découverte de l'OpenData. Il s'agit de parcourir un ensemble de sites proposant des données ouvertes.
- Choisir une (ou plusieurs) source(s) permettant de développer une application sur un thème donné choisi par vous. **Soyez créatifs !**
- Proposer un scénario d'application avec les jeux de données choisies.
- Examiner vos data : sont-elles directement « prêtes à l'emploi » ? Faut-il un pré-traitement particulier ?

NB : on ne développera pas l'application

Références

1. https://fr.wikipedia.org/wiki/Open_data
2. Le portail Région SUD des données ouvertes : <https://www.datasud.fr/>
3. Le portail français des données ouvertes : <https://www.data.gouv.fr/fr/>
4. Le site Etalab, avec des infos sur un hackaton : <https://www.etalab.gouv.fr/tag/hackathon>
5. Le portail américain des données ouvertes : <http://www.data.gov>
6. Pointeur vers des sites OD : <http://blog.visual.ly/data-sources/>
7. Transport : <https://data.sncf.com>
8. Le portail européen de données ouvertes
9. Exemples de projets opendata :
 - <http://askmedia.fr/blog/lopen-data-ca-sert-a-quoi-5-exemples-concrets/>
 - <http://data.worldbank.org/>

Autres outils

- csv2sql : <http://www.convertcsv.com/csv-to-sql.htm>
- json2csv : <http://www.convertcsv.com/json-to-csv.htm>

Qualité des données ouvertes

Exercice 1.2 – Qualité des données

Dans cet exercice, on s'intéresse à la qualité des données.

Travail à faire

- Parmi les sources de données que vous avez retenues à l'exercice 1.1, et en examinant ces données, pouvez-vous identifier les dimensions de qualité qui peuvent s'appliquer ?
 - a. Citer ces dimensions et illustrer les « problèmes de qualité » observés.
 - b. Ces sources de données sont-elles accompagnées d'un catalogue qui renseigne sur certaines dimensions de qualité ?
- Citer quelques outils de gestion de la qualité (recherche sur le Net).
- Tester un de ces outils avec un (des) jeu(x) de données retenu(s).