

# QUALITE DES DONNES

## 2021 - 2022

### TP n°1

#### Industrialiser un audit de qualité des données

### Contexte

Afin de faire face aux nouveaux défis induits par la mise en place des politiques de développement durable et de transition énergétique, une grande enseigne de location de véhicules a choisi de créer une filiale, **MOBILENVERT**, qui se propose de louer des véhicules électriques ou hybrides aux particuliers sensibilisés à ces enjeux environnementaux.

Cette offre a fortement séduit les clients désireux de se déplacer en zone urbaine car elle offre une alternative à la fois écologique et confortable aux offres de type **Vélib** ou **Autolib** actuellement proposées dans les grandes métropoles françaises.

Mais faute d'informations fiables et actualisées sur le parc de bornes de recharge des véhicules électriques, très peu de clients se hasardent à louer un tel véhicule pour des parcours inter cités. Et cette faible appétence des clients inquiète la direction de **MOBILENVERT** qui décide de se doter d'un logiciel capable de proposer une feuille de route sécurisée pour tout conducteur désirant se déplacer d'un point A à un point B (pas de risque de panne « sèche »).

Pour ce faire, une recherche est menée auprès de différents fournisseurs de données gratuite et payantes et **MOBILENVERT** a retenu le **fichier consolidé des Bornes de Recharge pour Véhicules Électriques** mis à disposition par les pouvoirs publics à l'adresse <https://www.data.gouv.fr/fr/datasets/fichier-consolide-des-bornes-de-recharge-pour-vehicules-electriques/>.

Echaudés par le passé dans l'emploi des fichiers proposés sur le site gouvernemental, il est décidé de réaliser un outil destiné à auditer la qualité des données contenues dans le fichier ciblé, réalisation qui vous est confiée par le DSI de l'entreprise.

## Travail demandé

L'objectif est de construire une application qui calcule automatiquement trois indicateurs de qualité sur toute version du fichier des Bornes de Recharge pour Véhicules Électriques publié régulièrement sur le portail [data.gouv](https://data.gouv.fr) :

1. Le % d'enregistrements corrects,
2. Le % d'enregistrements utilisables,
3. Le % d'enregistrements inutilisables.

La structure du référentiel (modèle de données) pouvant être amenée à évoluer, une grande flexibilité est exigée dans la gestion des règles syntaxiques et/ou sémantiques qui seront mises en œuvre sur les différents attributs du modèle de données.

Vu l'urgence de disposer d'une solution très rapidement, le DSI vous laisse carte blanche sur les choix de réalisation. Vous pouvez soit opter pour un développement spécifique et vous aurez alors à intégrer le besoin de flexibilité exprimé ci-dessus, ou vous pouvez construire votre application autour d'un des nombreux outils spécialisés du marché (panel représentatif disponible sur <https://www.capterra.fr/directory/31126/data-quality/software>).

## Planning du TP

- ☞ **15 mars 2022** : distribution du sujet (la composition des groupes - 4 étudiants maximum) est laissée à la libre appréciation des étudiants.
- ☞ **29 avril 2022** : présentation des résultats obtenus (aucun document n'est à rendre préalablement).