

Objectifs

L'objectif de ce projet consiste à récupérer des données ouvertes et d'être capable de les réintégrer dans une base de données.

De nombreuses entreprises se sont inscrites dans le mouvement Open Data, mettant en ligne de nombreuses données à destination du grand public.

- [opendatasoft](#)
- [10 sites de référence](#)

On trouvera notamment par exemple

- [l'INSEE](#) fournit de très nombreuses données (notamment une table répertoriant les naissances de 2011 (850.000 enregistrements) avec parmi les nombreux champs, l'âge de la mère, du père, le sexe, le nombre de naissances précédentes, les mois et jour de naissance, etc).
- L'institut Pasteur,
- [le gouvernement](#),
- les [affaires sociales](#),
- [l'INED](#),
- [Météo France](#),
- [La SNCF](#),
- la [RATP](#)
- [La Poste](#)
- [la MEL](#)
- [Météo-France](#)
- ou même [les fédérations de Football](#)

fournissent aussi de nombreuses données.

Soyez originaux ...

Toute la difficulté consiste à trouver des données relativement propres et suffisamment bien structurées (un CSV est évidemment bien plus préférable qu'un PDF) et pas trop volumineuse ($< 1Go$). Utilisez `/tmp` sur la machine local pour stocker temporairement ces données sans pénaliser votre quota.

1 A rendre

Un fichier script SQL contenant la marche à suivre à partir de la page de données pour importer ces données. Notamment :

- L'adresse WEB et les fichiers à télécharger
- Les scripts SQL à exécuter pour l'importation
- Quelques requêtes (10) intéressantes sur les tables ainsi importées (avec leur signification).