

## UCBL- TIW2- Interopérabilité 2021-2022

### TP Noté sur le Data Cleaning Nettoyage de données avec SQL

L'objectif de ce TP est de résoudre des problèmes de qualité de données en utilisant un langage de requête comme SQL.

Pour ce TP vous pouvez travailler sur la SGBD de votre choix (soit mySQL que vous avez installé lors du TP1 soit Postgres si vous avez des comptes au Département).

Vous allez utiliser les données ouvertes du [projet de loi de finances](#).

Ce TP est à rendre pour le dimanche 17 Octobre 2021 à 23:00 sur Tomuss (colonne **TP-Noté-1**) sous forme d'un rapport **PDF** contenant pour chaque type d'erreur:

- (1) une explication du problème de qualité et type d'erreur comme
  - violations de dépendances fonctionnelles;
  - violations de dépendances fonctionnelles conditionnelles;
  - problèmes de doublons;
  - problèmes de données manquantes

et une explication des moyens de le détecter et de le corriger;

- (2) une ou plusieurs requêtes SQL pour détecter ce type d'erreur;
- (3) une ou plusieurs requêtes SQL pour réparer ces erreurs.

Il faut trouver au moins 10 erreurs au total sur la table et compléter les étapes (1), (2) et (3) pour chaque erreur.

*Remarque:* il est possible d'utiliser des tables temporaires lors de la correction des erreurs.