

- **Jeu de données:**

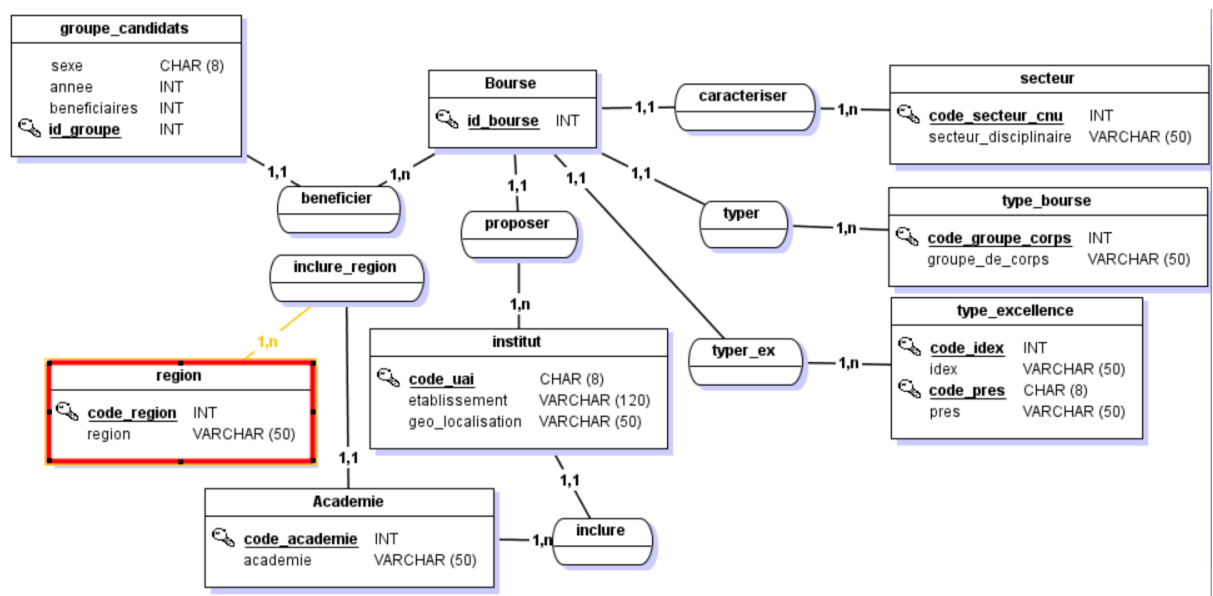
On a travaillé sur un jeu de données récupéré du site data.gouv.fr: "Les bénéficiaires de la prime d'excellence scientifique" sous forme de fichier csv.

Ce dernier présente les effectifs des bénéficiaires de la prime d'excellence corps et scientifique (ex prime d'encadrement doctoral et de recherche) organisé par sexe, secteur CNU et établissement.

On a opté pour la construction d'une base de données composée de 8 tables différentes qui respectent les règles de normalisation SQL:

groupe_candidats, Bourse, type_bourse, type_excellence, secteur, institut, academie, region.

Les tables ainsi que les cardinalités du MCD sont représentées dans la figure suivante:



L'ensemble des fichiers MCD et MLD, SQL peuvent être trouvés dans le rendu de projet.

- **Mode opérationnel:**

Le nettoyage ainsi que l'insertion des données des tables a été fait sous python (fichier *data_cleansing_and_insert.ipynb*). Le résultat du data cleansing est enregistré dans le fichier *bourse_vf.csv*.

Afin d'exploiter la Base de données, on commence par le lancement du fichier SQL *Bourses.sql*. Cela va créer la Base de données avec ses tables vides.

Ensuite on lance les fichiers SQL de chaque table nommé "Bourse_*.sql".

Les requêtes SQL ainsi que la data visualisation se trouve dans le notebook *Requete_sql_et_Data_visualisation.ipynb*.