**Conception de la base de données**

**Modèle conceptuel**

Il a fallu d’abord effectuer le modèle conceptuel de notre base et donc réfléchir à l’agencement et la pertinence des classes. Nous avons quatre types d’établissements avec des points communs à tous, des points communs uniquement entre les types d’établissements (entre lycées ou entre collèges) et des points communs uniquement entre les types d’élèves (filles/garçons). Nous avons décidé de créer une classe Etablissement qui regrouperait tous les points communs des quatre types d’établissements et quatre sous-classes à cette classe mère (collège garçons, collèges filles, lycées garçons et lycées filles). Nous avons fait le choix de sous-classes car nous voulions utiliser le code de l’établissement (présent dans nos tableaux) comme clé primaire de la classe mère Etablissement et donc de ses sous-classes qui en héritent mais aussi comme clé étrangère pour les relier à la classe mère Etablissement. Nous avons aussi créé des classes pour les entités géographiques des établissements : l’académie, le département, l’arrondissement et la ville.

*Mettre notre modèle conceptuel*

**Modèle logique relationnel**

Académie (ID\_académie)

Département (ID\_département, ID\_académie#)

Arrondissement (ID\_arrondissement, ID\_département#)

Ville/Chef-lieu (ID\_ville, ID\_arrondissement#)

Etablissement (ID\_etablissement, Date\_fondation, Type\_etablissement, Type\_eleves, Superficie\_totale, Superficie\_libre, Superficie\_bâtie, Recette\_totale, Nbre\_total\_eleves, etc., ID\_ville#)

Lycée garçons (ID\_etablissement, comparaison\_recettes\_depenses, depenses\_personnel, depenses\_materiel, rentes, etc.)

Lycée filles (ID\_etablissement, depenses\_personnel, depenses\_materiel, comparaison\_recettes\_depenses, etc.)

Collège garçons (ID\_etablissement, mobilier, recettes\_frais\_pension, depenses\_personnel, depenses\_materiel, nbre\_prof, etc.)

Collège filles (ID\_etablissement, etat\_mobilier, recettes\_ordinaires, depenses\_ordinaires, etc.)

**Modèle physique**

Pour créer les tables et leurs attributs, il a d’abord fallu repérer les points commues au niveau des variables des établissements. Mais, nos jeux de données, bien que comportant de nombreuses variables communes, n’avaient aucune harmonie entre eux au niveau des intitulés des variables. Ainsi, nous pouvions avoir quatre variables désignant les recettes totales, mais avec des intitulés tous différents. Cela a donc rendu la recherche des variables communes très longue et pénible. Un travail pour renommer les variables a donc été fait sur R. Une fois toutes les variables communes repérées, nous avons pu commencer à créer notre base de données dans Mariadb. Nous avons créé nos tables et leurs relations. Nous avons complété la table Etablissement avec les variables communes trouvées aux quatre jeux de données. Puis les sous-classes avec les variables communes aux filles, aux garçons, aux lycées et aux collèges.

Pour importer les données à partir de fichiers csv, pour ne pas à avoir à rentrer tout à la main, nous avons créé autant de fichiers csv que de tables. Pour les tables Academie, Département, Arrondissement et Ville, il a fallu constituer des tableaux contenant tous les codes que les établissements pouvaient avoir, et associer ensuite la bonne valeur de clé étrangère à la bonne valeur de clé primaire. Par ailleurs, comme les codes des arrondissements et des villes étaient compris entre 1 et 6, et que ces codes étaient supposés être nos clés primaires, donc uniques, il a fallu récréer des codes à partir des codes des entités géographiques supérieures (par exemple, code arrondissement = code département et code arrondissement). Une fois ces ajustements effectués, nous avons pu importer les données à partir de fichiers csv. Nous avons aussi décidé d’ajouter les noms des départements et villes, mais il a fallu associer les codes de nos fichiers aux intitulés correspondants, pas présents dans nos tableaux, mais indiqués dans des fichiers PDF sur la page web de description de nos jeux de données (<https://journals.openedition.org/acrh/2891?lang=fr>), ce qui a également demandé du temps. Pour les tables Etablissement, Collège fille, Collèges garçons, Lycées filles et Lycées garçons, il a également fallu retravailler les fichiers de base. Nous avons sélectionné les variables qui correspondaient aux attributs choisis pour la base de données. Nous avons eu un problème au niveau des valeurs manquantes, d’abord recodées en NA, car nous ne pouvions pas importer ces données dans les tables avec des attributs définis comme numériques. Nous avons donc finalement laissé les données manquantes tels quelles et au moment de l’importation elles ont été remplacées par des valeurs nulles.