

Partie 1 - Implantation du schéma de la base de données

Introduction au projet

L'objectif de ce projet est vous faire implanter, peupler puis exploiter une base de données réaliste. Nous allons nous appuyer sur vos compétences de traduction de modèles de données (diagramme de classe UML) et de définition d'un schéma relationnel en SQL, mais aussi de Mathématiques pour exploiter des statistiques tirées de cette base de données, de gestion de projet pour la planification, et d'Anglais pour la présentation de votre travail.

Le sujet de cette année porte sur des informations à propos des effectifs, des taux de réussite, et des "indices de position sociale" et de la valeur ajoutée des établissements de l'enseignement primaire et surtout secondaire en France sur plusieurs années.

Pour cela nous vous fournissons dans un premier temps un diagramme de classe à traduire dans le modèle relationnel. En complément de ce diagramme de classe, nous vous fournissons des données présentées sous forme de plusieurs fichiers CSV disponible en *open data*. Pour simplifier, nous avons restreint le diagramme de classe et vous constaterez que les fichiers CSV contiennent plus d'informations. Pour la partie statistiques, nous vous fournirons une base complète et peuplée, accompagnée d'un diagramme de classe un peu plus complexe.

Présentation du modèle de données

Comme indiqué dans l'introduction, nous souhaitons modéliser une partie des données concernant les établissements de l'enseignement primaire et secondaire, en particulier les collèges.

Pour cela nous vous proposons le diagramme de classe dans la figure 1. Des fichiers CSV sont également disponibles pour vous donner une idée de la nature et de la taille des attributs par exemple.

Le diagramme de classe représente tout d'abord les informations de situation géographique des établissements. Un établissement est localisé dans une commune, qui elle-même se situe dans un département, inclus dans une région. L'établissement fait partie également d'une académie qui est un découpage du ministère de l'Education Nationale. Ces académies ne suivent pas forcément le découpage en régions. Un établissement possède un type (collège, lycée par exemple) et peut se situer à proximité d'un quartier prioritaire.

Ensuite, les données concernant tous les établissements (association `caracteristiques_tout_etablissement`), puis uniquement les collèges (classes-associations `caracteristiques_college` et `caracteristiques_selon_classe`) sont modélisées.

Certains attributs auront des valeurs fixées qui ne sont pas données de façon évidente dans le CSV, et en particulier qui pourront revêtir plusieurs formes selon le fichier CSV. Vous pourrez choisir celle qui vous semble la plus complète (avec le plus d'information).

Ce dont vous disposez pour travailler

- Le présent énoncé `sae204_2024_partie1.pdf`.
- Les fichiers CSV : toutes les informations de ces fichiers ne sont pas utilisées dans le diagramme de classe qui est simplifié.
 - `fr-en-college-effectifs-niveau-sexe-lv.csv` : en plus de quelques informations géographiques, effectifs dans des établissements de différents types (segpa, ulis, filles/garçons, langues vivantes étudiées);
 - `fr-en-etablissements-ep.csv` : informations sur les établissements en zone d'éducation prioritaire (EP);
 - `fr-en-indicateurs-valeur-ajoutee-colleges.csv` : indicateurs de "performance" des collèges;
 - `fr-en-ips-colleges-ap2022.csv` : indices de position sociale des collèges en 2022-2023.
- un fichier tableur nommé `dico_data.xlsx` décrivant le dictionnaire des données pour vous aider à la compréhension du diagramme. Ont été écartées les données calculables qui ne doivent évidemment pas être stockées. Vous pourrez réutiliser les noms de la première colonne qui sont ceux utilisés dans le diagramme de classe (utile pour copier-coller les noms d'attributs pour votre schéma relationnel en SQL).
- le diagramme de classe sous forme d'un fichier image séparé du présent énoncé, sous le nom de `sae204_statistiques_colleges.jpg`.

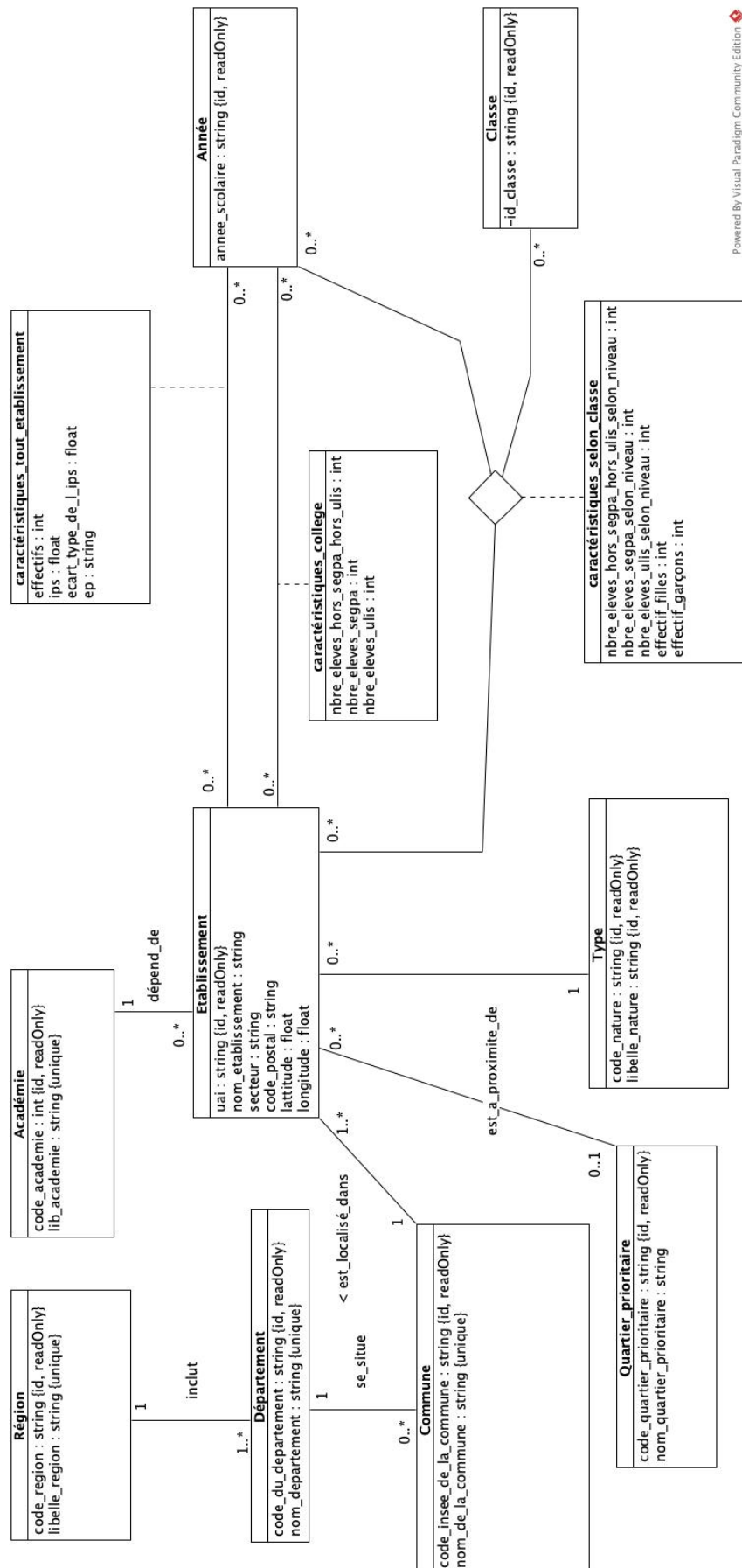


FIGURE 1 – Données dans l’enseignement primaire et secondaire en France - Diagramme de classe

Mise en œuvre et travail demandé

- **Réfléchir au schéma relationnel** : graphe des contraintes d'intégrité référentielles, traduction du diagramme de classes sous forme d'un script SQL de **création de bout en bout** de la structure de la base de données. Ce script sera à déposer sur Moodle dans un dépôt dédié créé pour l'occasion.
- **Implantation sur votre base** commune nommée `pg_sae204_k12` par exemple, si vous êtes le deuxième binôme du groupe K1. Pour cela il vous faut constituer des binômes sur [ce fichier Excel qui est en ligne](#). Cette implantation consiste à traduire le diagramme de classe en un schéma relationnel exprimé en SQL.

Les consignes suivantes devront **obligatoirement être respectées** :

- L'implantation du schéma relationnel de la base de données sera réalisée avec le SGBD PostgreSQL à l'aide d'un script SQL que vous nommerez `statistiques_colleges_db_create.sql`. Ce script est supposé pour s'exécuter de bout en bout et réinitialiser le schéma et définir sa structure.
- Les noms des schémas seront définis en minuscules.
- Les relations (tables) seront définies au sein d'un schéma nommé `colleges`.
- Les noms des relations de base seront définis en minuscules et préfixés par `_`. Exemples : `_departement`, `_region`.
- Les contraintes seront nommées explicitement.
- les noms d'attributs seront définis en minuscules et reproduiront les variables d'instance ou noms de rôle (au singulier) du diagramme de classes.

Livrables attendus pour cette première partie (date de rendu : 24/05/2023 à 18h00 au plus tard, à adapter selon les groupes) :

- un fichier `statistiques_colleges_db_create.sql` à déposer sur l'ENT
- l'implantation de la structure de la base à l'aide de ce script dans un schéma nommé `colleges` sur votre base commune `pg_sae204_k12` (si vous êtes membre du groupe K12).