



# Projet Go Explore : “Sur la route des Artistes”

...

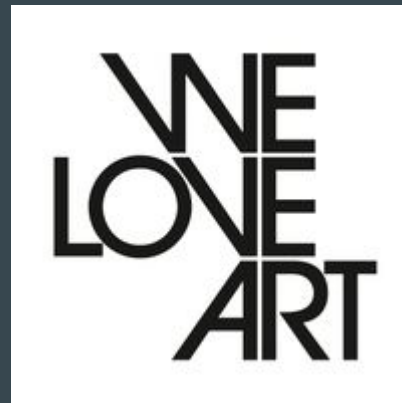
Maëlle Coriou



# SOMMAIRE

- 1- Le projet
- 2- Les données sources
- 3- Schéma fonctionnel / logiciel
- 4- Sélection des données
- 5- Schémas MCD et MLD
- 6- Nettoyage des données : Ex. Artistes
- 7- Sauvegardes
- 8- Création Bdd socle PostgreSql
- 9- Insertion des données
- 10- Optimisation d'une requête
- 11- Sur la route de Léonard de Vinci
- 12- Rétrospective : Actions à mener

# 1- Le Projet



# 1- Le Projet, Objectifs

Cibler les lieux d'expositions :

recherches par Artistes, titre d'œuvres, techniques, périodes

Analyse par musées :

du nombre d'artistes

du nombre d'œuvres par époque

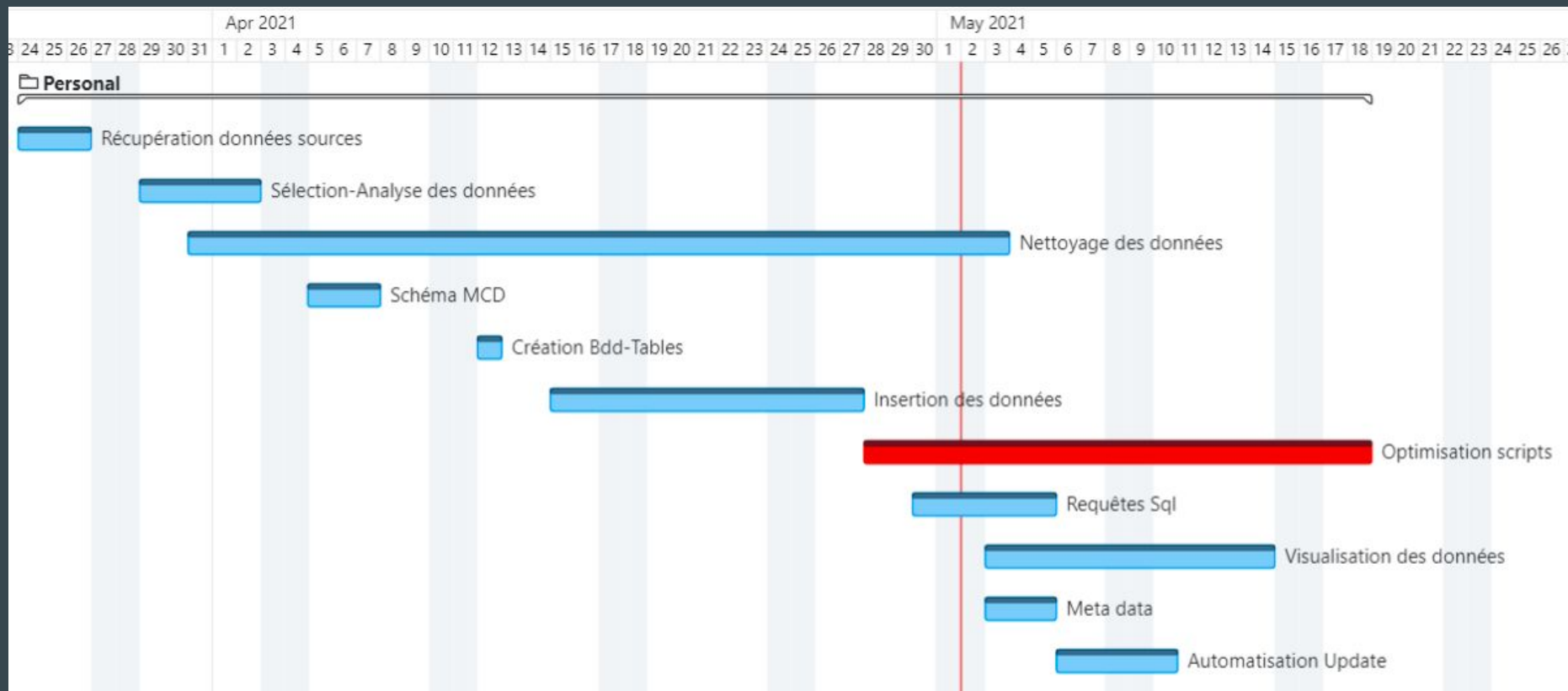
du nombre d'œuvres par technique et matériaux

du nombre d'œuvres par domaine

Analyse par région :

domaines principaux, artistes phares, époques, nombre de musées, distance

# 1- Le Projet : Planning



## 2- Les données sources



**Base Joconde** - Inventaire des musées de France

650k Enregistrements,

Upload format Csv, Json ou Api

[Collections des musées de France : extrait de la base Joconde](#)

**Base Muséofile** - Liste et infos des musées de France

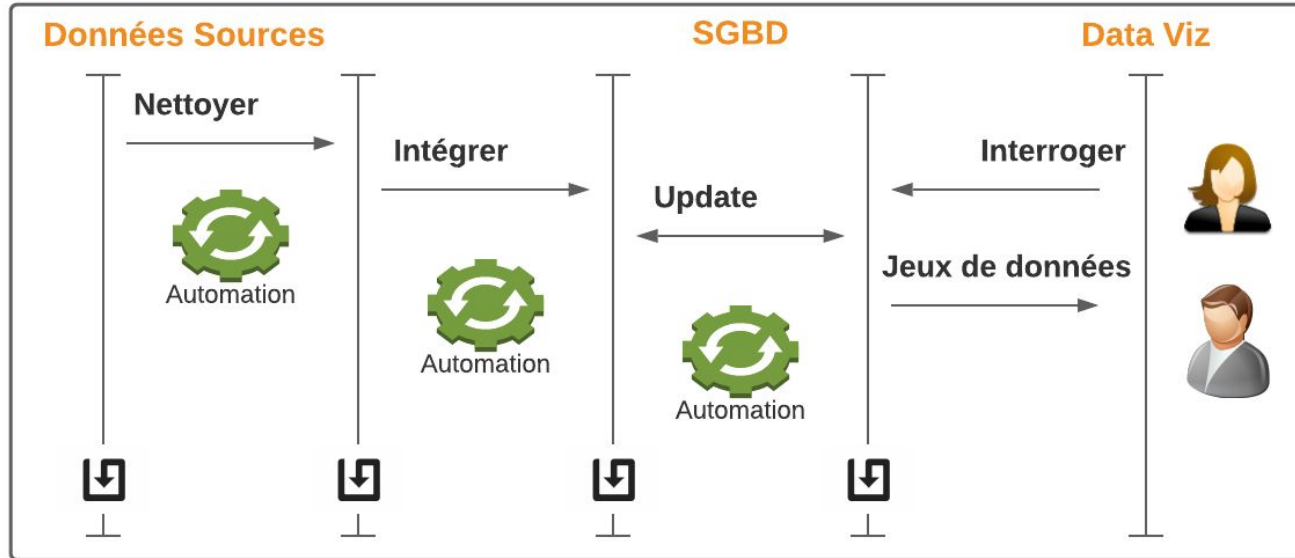
1.220 Enregistrements,

Upload format Csv, Json ou Api

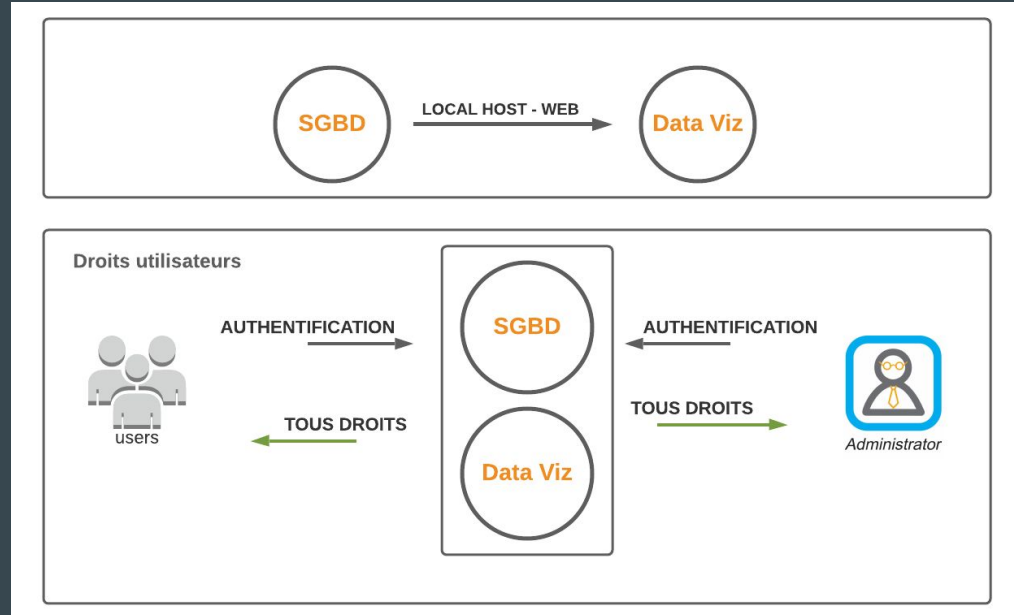
[Musées de France : base Muséofile — Ministère de la Culture](#)

Licence ouverte : [Licence Ouverte Version 2.0 \(etalab.gouv.fr\)](#)

### 3. Schéma fonctionnel

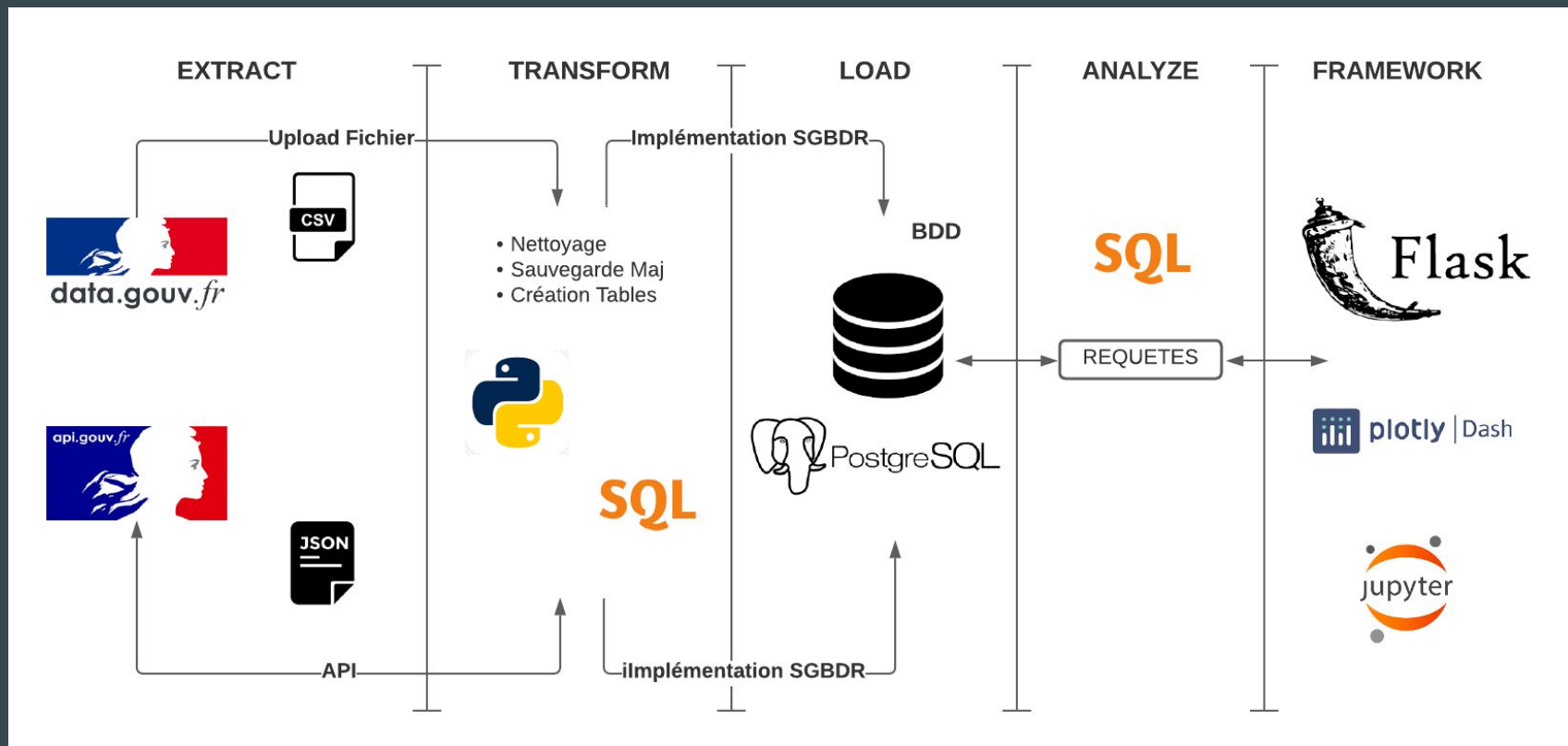


### 3. Schéma communication





### 3- Modélisation Logicielle du projet



## 4- Sélection des données : base Joconde

17 champs sur 40 :

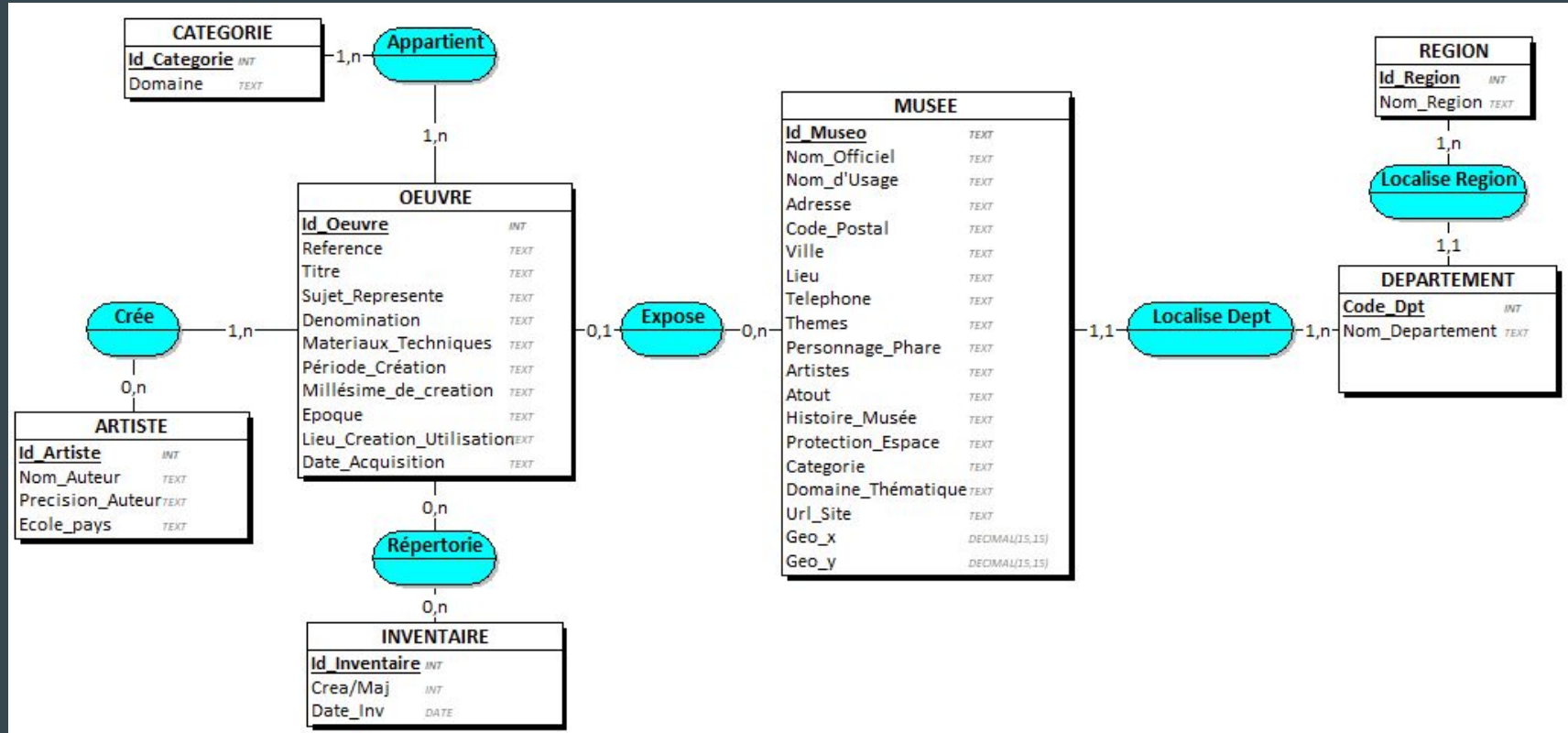
Nom Champ	Type	Exemple information
Référence	texte	M5037010624
Auteur	texte	DANTAN Jean Pierre;DANTAN le Jeune (dit)
Date acquisition	texte	1888 vers
Dénomination	texte	buste
Date de mise à jour	date	
Date de création	date	2001-02-22
Domaine	texte	sculpture
Ecole-pays	texte	France
Epoque	texte	
Lieu de création/utilisation	texte	Italie;Rome (lieu d'exécution)
Millésime de création	texte	1829
Identifiant Museofile	texte	M5031
Précision auteur	texte	Paris, 1800 ; Bade, 1869
Période de création	texte	2e quart 19e siècle
Sujet représenté	texte	portrait charge (Vernet Horace, homme, en buste, satirique, peintre, dessinateur)
Matériaux – techniques	texte	bronze
Titre	texte	Carle Vernet, Charles-Horace, dit (Bordeaux, 1758-Paris, 1836)

# 4- Sélection des données : base Museofile

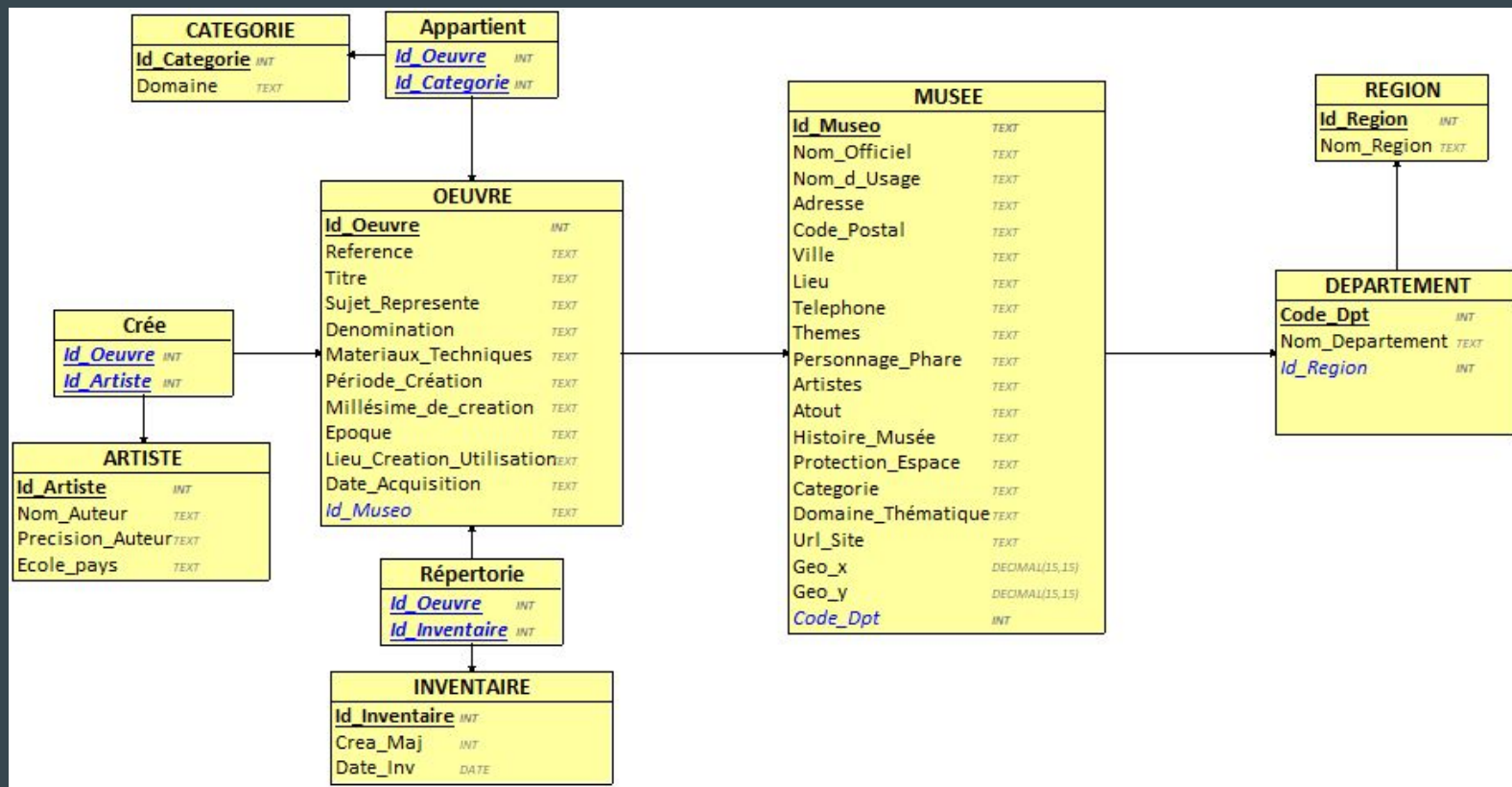
23 champs :

Nom Champ	Type	Exemple information
Identifiant	texte	M0144
Adresse	texte	12, rue Camille Rodier
Artiste	texte	Théodore Lévigne (1848 - 1912) ; Paul Jean Baptiste Gasq (1860 – 1944)...
Atout	texte	Collections archéologiques gallo-romaines (site des Bolards)....
Catégorie	texte	Musée de site
Code Postal	texte	21700
Domaine thématique	texte	Archéologie;Beaux-arts;Ethnologie;Histoire
Département	texte	Côte-d'Or
Date de saisie	texte	2019-01-21
Histoire	texte	Le docteur Duret (1794-1874), maire de la ville, jette les bases du premier musée...
Intérêt	texte	Ancienne maison de vins appartenant à la famille de Camille Rodier ...
Lieu	texte	Ancienne Maison de Vins
Nom officiel	texte	musée municipal
Nom usage	texte	Musée de Nuits-Saint-Georges
Personnage phare	texte	
Protection bâtiment	texte	
Protection espace	texte	Aux abords d'un monument historique
Région	texte	Bourgogne-Franche-Comté
Téléphone	texte	03 80 62 01 37
Thèmes	texte	Archéologie nationale : Gallo-romain, Paléo-chrétien (Mérovingien);Beaux-Arts...
URL	texte	<a href="http://www.ot-nuits-st-georges.fr/fr/fiche.html?id_fiche=2437">www.ot-nuits-st-georges.fr/fr/fiche.html?id_fiche=2437</a>
Ville	texte	Nuits-Saint-Georges
geolocalisation	geo_point_2d	[47.138661,4.950405]

# 5- Modélisation retenue Bdd Socle : MCD Entités / Associations



# 5- Modélisation retenue Bdd Socle : MLD





An abstract painting with vibrant, bold colors including yellow, green, blue, red, and purple. It features thick, expressive brushstrokes and a central figure that appears to be a face or a head, rendered in a stylized, almost cubist manner.

**LOVE  
THE ART**

A black and white photograph of Winston Churchill. He is shown from the chest up, looking slightly to the right with a serious expression. He is wearing a dark suit and a white shirt. The background is out of focus, showing what appears to be a large, domed structure, possibly a dome of a building.

**HATE  
THE ARTIST**

## 6- Nettoyage des données - Les Artistes : Exemples

Problématiques du champ Auteur :

ALIGNY ThÃ©odore;CARUELLE D'ALIGNY(dit) || VUE PRISE DANS L'ILE DE CAPRI

ABC nom_artiste	ABC titre_oeuvre
CARUELLE D'ALIGNY	VUE PRISE DANS L'ILE DE CAPRI
ALIGNY THEODORE	VUE PRISE DANS L'ILE DE CAPRI

## 6- Nettoyage des données - Les Artistes : Exemples

Angelo di COSIMO (peintre);BRONZINO IL (dit) || Portrait de la noble dame Dianora Frescobaldi

ABC nom_artiste	ABC titre_oeuvre
ANGELO DI COSIMO	PORTRAIT DE LA NOBLE DAME DIANORA FRESCOBALDI
BRONZINO IL	PORTRAIT DE LA NOBLE DAME DIANORA FRESCOBALDI



## 7- Sauvegardes : Cycle de vie des fichiers

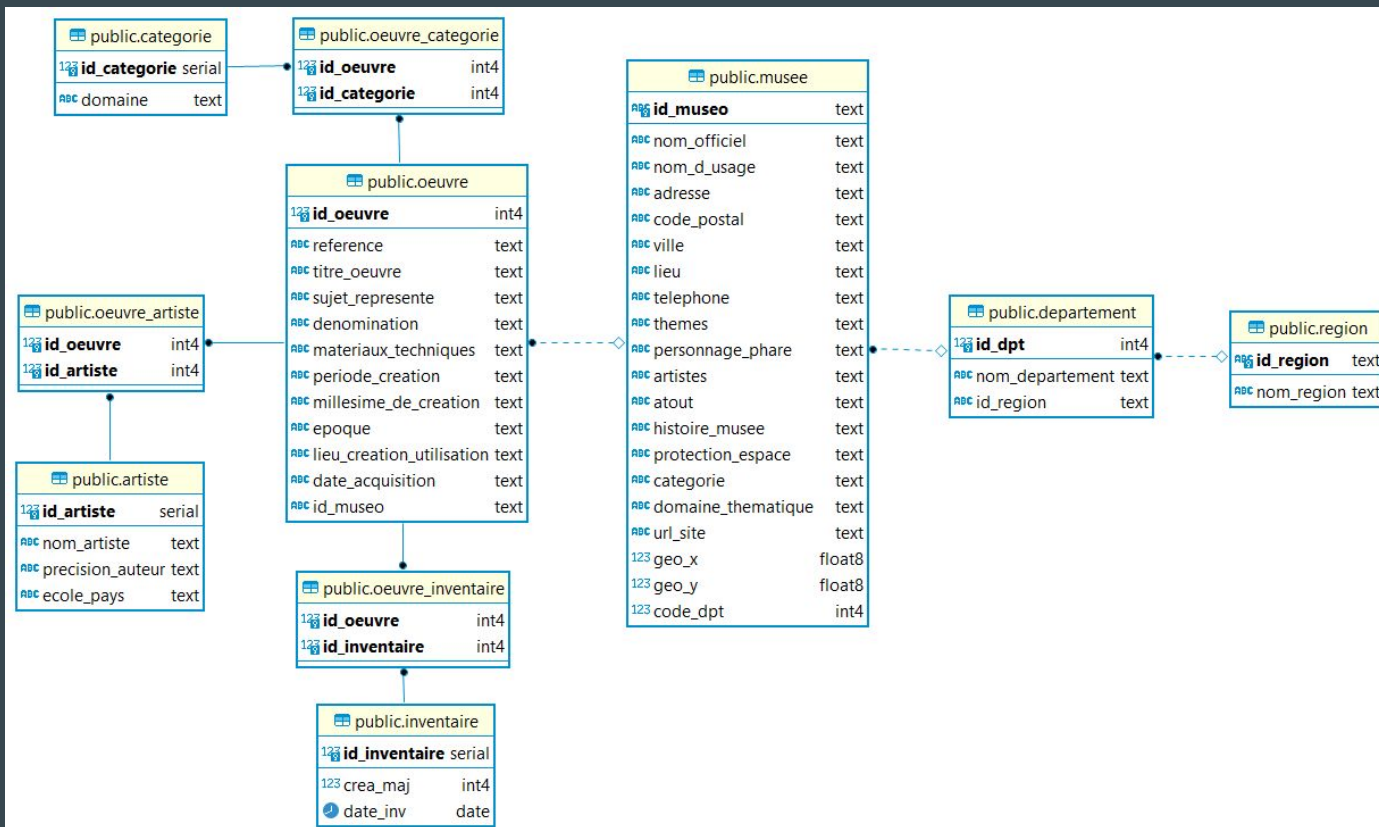
```
-- Data
| -- Csv_Bdd
|   | -- Table_Auteurs.csv
|   | -- Table_Departements.csv
|   | -- Table_Domaines.csv
|   | -- Table_Inventaire.csv
|   | -- Table_Musees.csv
|   | -- Table_Oeuvres.csv
|   | -- Table_Regions.csv
| -- Maj
|   | -- Check_Dpt_Region.csv
|   | -- Departements.csv
|   | -- Dpt_Region_Code_Region.csv
|   | -- Inventaire_V2.csv
|   | -- Pivot_Auteurs.csv
|   | -- Pivot_DateCrea.csv
|   | -- Pivot_Domaines.csv
|   | -- Table_Region.csv
|   | -- Update_check_V2.csv
| -- Sources
|   | -- Region.csv
|   | -- base-joconde-arts-decoratifs.csv
|   | -- base-joconde-peintures.csv
|   | -- base-joconde-sculptures.csv
|   | -- base_museofile.csv
```

# 7- Sauvegardes : Métadonnées des fichiers

The screenshot shows the MongoDB Compass web interface. The browser tab is titled 'meta\_bdd\_V3.meta\_rep... Documents'. The main header displays the collection name 'meta\_bdd\_V3.meta\_repertoire\_V3' and statistics: 69 documents (45.2KB total size, 671B avg size) and 1 index (36.0KB total size, 36.0KB avg size). The 'Documents' tab is selected, showing a filter bar with the query '{ field: 'value' }'. Below the filter bar, there are buttons for 'ADD DATA', 'VIEW' (with icons for list, JSON, and grid), and a status bar indicating 'Displaying documents 61 - 69 of 69' with navigation arrows and a 'REFRESH' button. The main content area displays the raw JSON document for the selected file.

```
{
  "_id": ObjectId("60a0cf3abff7e014d677baab"),
  "ExifTool Version Number": "12.25",
  "File Name": "base-joconde-arts-decoratifs.csv",
  "Directory": "C:/Users/utilisateur/SIMPLON/DEV-DATA/projet-chef-d-oeuvre/Version_3/D...",
  "File Size": "6.4 MiB",
  "File Modification Date...": "2021:05:14 11:40:19+02:00",
  "File Access Date/Time": "2021:05:16 09:52:26+02:00",
  "File Creation Date/Time": "2021:05:14 18:36:25+02:00",
  "File Permissions": "-rw-rw-rw-",
  "File Type": "CSV",
  "File Type Extension": "csv",
  "MIME Type": "text/csv",
  "MIME Encoding": "utf-8",
  "Byte Order Mark": "No",
  "Newlines": "Windows CRLF",
  "Warning": "[Minor] Not counting rows past 1000",
  "Delimiter": "Semicolon",
  "Quoting": "Double quotes",
  "Column Count": "40"
}
```

# 8- Création Bdd : Postgresql : Schéma DBeaver



# 9- Insertion des données Bdd Socle : Récap PG Admin

Insertion des données Table Oeuvre :

Art-décoratif : 6.080 rows  
Céramique : 27.631 rows  
Peinture : 69.826 rows  
Sculpture : 33.548 rows  
Photographie : 33.435 rows  
Dessin : 243.020 rows

**Total à insérer = 413.540**

Table name	Tuples inserted
artiste	29,023
categorie	232
departement	102
inventaire	3,561
musee	1,221
oeuvre	413,540
oeuvre_artiste	479,732
oeuvre_categorie	527,473
oeuvre_inventaire	475,176
region	15

# 10- Optimisation d'une performance : Insertion Update Api

```
response = requests.get('https://data.culture.gouv.fr/api/records/1.0/search/?  
dataset=base-joconde-extrait&q=&facet=domaine')
```

```
data = []  
response_1 = json.loads(response.text)  
  
for i in range(len(response_1['facet_groups'][0]['facets'])):  
    data.append([response_1['facet_groups'][0]['facets'][i]['name']])
```

```
# Liste Data sous df, formater en minuscule pour recherche doublons Bdd  
Domaine_Api = pd.DataFrame(data, columns=['domaine'])  
Domaine_Api.domaine = Domaine_Api.domaine.str.lower()  
# Remplacer les caractères Iso non traduits en UTF-8  
Domaine_Api = Domaine_Api.replace(to_replace=("Ã©", "Ã-", "Ã'", "Ã", "Ã´", "Ã§", "Ãª", "Ã¹", "Ã¼",  
"Ã«", "Å", "à§", "ãª", "à%", "à^", "àª"), value=("é", "ï", "è", "à", "ô", "ç", "ê", "ù", "ü", "ë",  
"œ", "ç", "ê", "É", "ê", "ê"), regex=True)
```

## 10- Optimisation d'une performance : Insertion Update Api

```
Insertion_Donnees = '''
    INSERT INTO CATEGORIE
    (Domaine)
    VALUES (%(domaine)s);
'''

functions_PgSql.inserer_df(conn, Insertion_Donnees, Domaine)
```

## 10- Optimisation d'une performance : Insertion Update Api

```
requete = """
    SELECT c.domaine,
           COUNT(c.domaine) AS Nb_Domaine
    FROM categorie c
    GROUP BY c.domaine
    ORDER BY Nb_Domaine DESC;
"""
Categories_check = pd.read_sql_query(requete, conn)
```

(114, 2)

	domaine	nb_domaine
0	balnéaire	2
1	temps modernes	2
2	arts graphiques	2
3	art religieux	2

## 10- Optimisation d'une performance : Insertion Update Api

```
# Suppression des éventuels doublons de la table Catégorie
# Supprimer le dernier Id créé
suppression_doublons = ''' DELETE
                           FROM categorie
                           WHERE id_categorie IN (
                               SELECT MAX(id_categorie)
                               FROM categorie
                               GROUP BY domaine
                               HAVING COUNT(domaine) > 1)'''
```

▶ M↓

```
functions_PgSql.modification_table(conn, suppression_doublons)
```



# 10- Optimisation d'une performance : Insertion Update Api

Optimisation :

```
# Insertion des données, ON CONFLICT sur la contrainte unique domaine pass  
# Evite les doublons  
Insertion_Donnees = '''  
    INSERT INTO CATEGORIE  
    (Domaine)  
    VALUES (%(domaine)s)  
    ON CONFLICT (domaine)  
    DO NOTHING;  
'''
```

# 10- Optimisation d'une performance : Insertion Update Api

Insertion Api V1 :

--- 4.82623028755188 seconds ---

Insertion Api V2 :

--- 1.8876831531524658 seconds ---

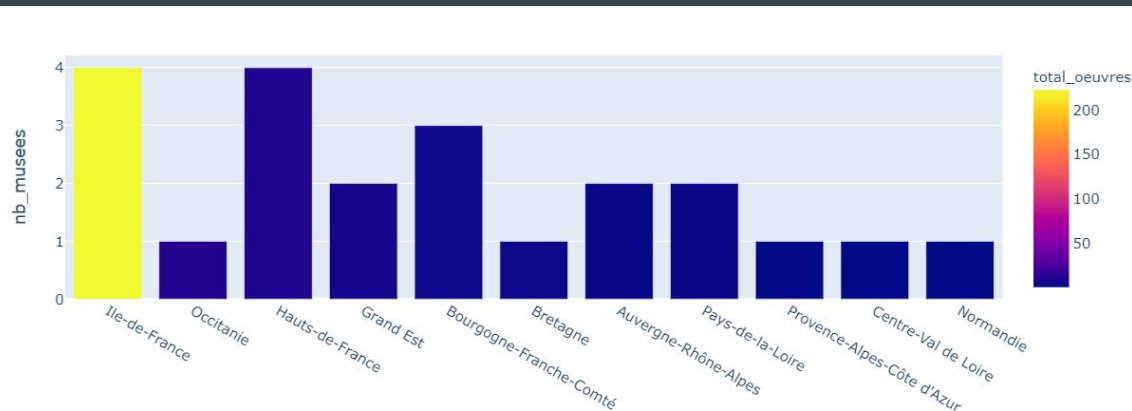
Table name	Tuples inserted	Tuples updated	Tuples deleted	Tuples HOT updated	Live tuples
artiste	14,985	0	0	0	14,985
categorie	130	0	16	0	114
departement	102	0	0	0	102
inventaire	1,799	0	0	0	1,799
musee	1,218	0	0	0	1,218
oeuvre	69,678	0	0	0	69,678
oeuvre_artiste	81,055	0	0	0	81,089
oeuvre_categorie	76,939	0	0	0	76,993
oeuvre_inventaire	78,367	0	0	0	78,498
region	15	0	0	0	15

Table name	Tuples inserted	Tuples updated	Tuples deleted
artiste	14,985	0	0
categorie	114	0	0
departement	102	0	0
inventaire	1,799	0	0
musee	1,218	0	0
oeuvre	69,678	0	0
oeuvre_artiste	81,129	0	0
oeuvre_categorie	76,940	0	0
oeuvre_inventaire	78,367	0	0
region	15	0	0

# 11- Sur la route de Léonard de Vinci

(38, 4)

	nom_artiste	nb_oeuvres	nom_musee	ville
0	VINCI LEONARDO DA	187	musée du Louvre	Paris
1	VINCIDOR TOMMASO	12	musée du Louvre	Paris
2	LEONARDO DI SER PIERO DA VINCI	10	musée du Louvre	Paris
3	VINCI LEONARD DE	8	musée du Louvre	Paris
4	LEONARD DE VINCI	7	musée Gustave Moreau	Paris
5	LEONARD DE VINCI	6	musée Ingres	Montauban
6	LEONARDO DI SER PIERO DA VINCI	5	musée Ingres	Montauban
7	VINCI LEONARD DE	5	musée Ingres	Montauban
8	LEONARD DE VINCI	4	musée des Beaux-Arts	Strasbourg
9	LEONARDO DI SER PIERO DA VINCI	3	musée des Beaux-Arts	Rennes
10	LEONARD DE VINCI	3	Musée Condé	Chantilly
11	LEONARD DE VINCI	2	musée de la Chartreuse	Douai
12	LEONARD DE VINCI	2	musée Boucher de Perthes	Abbeville





# 12- Rétrospective : Les 4 L

## LIKED

Le projet

Apprendre

Une Bdd créée !

Résultat

## LEARNED

Il y a toujours une solution

Tellement de choses !

## LAKED

Travail d'équipe

Expérience

## LONGED

Que mon projet vous a plût !

## 12- Rétrospective : Les Actions

- Script calcul la distance entre chaque musée.
- Crontab pour le backup plutôt que schedule.
- Scraping
- Développer une interface IHM
- Optimiser le script

# Liens ressources

Photos :

[Léonard l'expo-pulaire](#)

[Matisse sculpte la couleur avec des ciseaux : les découpages à la Tate Modern](#)

[Livre illustré de Alexander Calder, J. Prévert : FÊTES. 7 aquatintes originales de Calder \(1971\) sur Amorosart](#)

Code:

[Petit cours de modélisation - C.Darmangeat](#)

[Manipulation des données avec Pandas](#)

[Guide des expressions régulières — Documentation Python 3.9.2](#)

[Filtrer les données manquantes \(NaN, NULL\) d'une DataFrame avec Pandas ?](#)

[5 Must-Know Pandas Operations on Strings | by Soner Yıldırım](#)

[Splitting Columns With Pandas](#)

[pandas.wide\\_to\\_long — pandas 1.2.4 documentation](#)