



DÉFINITION ET DESCRIPTION DES SOURCES D'INFORMATION

PROJET D'INTEROPÉRABILITÉ

GROUPE 4

IDRISS BENGUEZZOU
INESS BOUABID
ALI BOUGASSAA
GHILAS MEZIANE

08/02/2023

Table des matières

1	Objet du document	2
2	Description des sources des données	2
2.1	Liste et localisation des musées en France	2
2.2	Informations sur le patrimoine culturel en France	2
2.3	Informations sur la fréquentation des musées en France	3
2.4	Information sur le patrimoine mondial en France	3

1 Objet du document

L'objectif du document est de fournir une définition et une description détaillées des sources d'informations qui seront utilisées pour alimenter la base de données centralisée du projet de promotion de la vie culturelle dans la région.

Pour chaque source d'information, le document va identifier et décrire les types d'informations que la source fournit, le format dans lequel les informations sont disponibles, les moyens d'accéder à ces informations, ainsi que les méthodes et les outils utilisés pour extraire, nettoyer et stocker les données dans la base de données centrale.

2 Description des sources des données

2.1 Liste et localisation des musées en France

Le site **<https://data.culture.gouv.fr/>** contient des informations sur la liste et la localisation des musées de France. Il contient plusieurs colonnes mais celles qui nous importe le plus sont les suivantes :

- Le nom du musée.
- L'adresse.
- Le code postale.
- La ville.
- La latitude.
- La longitude.

Le jeu de données vient dans différents formats : des données au formats CSV, JSON, et des feuilles de calculs Excel.

2.2 Informations sur le patrimoine culturel en France

Le site **<https://tourisme62.opendatasoft.com/>** nous fourni des données sur le patrimoine culturel dans le département du Pas-de-Calais en France.

Chaque instance de données comprend les informations suivantes :

- Le nom du patrimoine culturel.
- Le type du patrimoine (site archéologique, monument historique...etc).
- La commune.
- Le code postal.
- L'adresse.
- La latitude.
- La longitude.

Le jeu de données vient dans 3 formats différents : JSON, CSV et Feuille de calcul Excel.

2.3 Informations sur la fréquentation des musées en France

Chaque ligne du jeu de données du site <https://data.opendatasoft.com/> représente un musée et ses données de fréquentation pour une seule année, il contient aussi les informations sur les visiteurs.

Chaque ligne comporte les champs suivants :

- Le nom du musée.
- L'année.
- La fréquentation, soit le nombre de visiteur du musée selon l'année.
- L'adresse.
- La ville.
- Le code postal.

Les données sont représentées dans 3 formats : CSV, JSON et Excel.

2.4 Information sur le patrimoine mondial en France

Le site <https://whc.unesco.org/> contient les informations sur le patrimoine mondiale de l'UNESCO. Pour l'alimentation de la base de connaissances de Wikibase nous allons utiliser uniquement les données du patrimoine présent en France.

Le jeu de données est structures dans un seul format ; XML et contient les informations détaillées sur chaque site du patrimoine. Chaque entrée, présente comme éléments :

- Le nom du site.
- Le pays ou se trouve le site.
- La date d'inscription sur la liste du patrimoine mondiale.
- Le statut de l'inscription.
- Les raisons pour lesquelles le site a été inclus dans la liste.
- La description du site.
- Les coordonnées géographiques.