

# Projet de Scraper centres de tri FFB

Le projet est une application complète permettant de collecter, organiser et exploiter les données des centres de tri de déchets de chantier référencés sur le site de la FFB. Il combine une interface graphique intuitive avec des fonctionnalités avancées de web scraping et d'export de données.

## 1- Objectif du Projet

Créer un **outil automatisé** avec interface graphique pour :

- Scraper les **centres de tri de déchets de chantier** depuis le site de la FFB.
- Traiter différents **types de déchets** (classés en 11 groupes).
- Offrir des **options d'export** (CSV, JSON, Word, Excel, Carte interactive).
- Permettre une **utilisation facile et modulaire** pour différents profils utilisateurs (recherche simple, multi-catégorie ou avancée).

**[Difficultés rencontrées et solutions apportées \(cliquez ici\)](#)**

**[Arbre de décision des types de déchets \(cliquez ici\)](#)**

## 2- Fonctionnalités Principales

### 1. Collecte Automatisée des Données

- **Navigation automatisée** via Selenium pour interagir avec le site FFB
- **Extraction des données** des centres de tri (nom, numéro, email, adresse, coordonnées, types de déchets acceptés, etc.)
- **Gestion de plusieurs modes de recherche:**
  - Recherche multi-catégories (scan complet)
  - Recherche par catégorie spécifique
  - Recherche par nouvelles catégories

### 2. Stockage des Données

- **Base de données SQLite** pour persister les résultats
- **Détection de doublons** pour éviter les répétitions
- **Historique des recherches** avec dates de scraping

### 3. Interface Utilisateur

- **Interface graphique TKinter** intuitive
- **Visualisation des résultats** dans un tableau interactif
- **Fenêtre de détails** pour chaque centre
- **Options d'export** multiples (Word, Excel, CSV, JSON)
- **Génération de cartes** avec Folium

#### 4. Export des Résultats

- Formats supportés: Word, Excel, CSV, JSON
- **Cartographie interactive** des centres
- **Mise en page professionnelle** dans les documents Word/Excel

### 3- Architecture du Code

Le projet est structuré en 3 fichiers principaux:

1. **Selenium\_auto\_Simple.py**: Version simplifiée du scraper par catégorie unique ou par type de plateforme ou type de déchet.
2. **Selenium\_auto\_Complet.py**: Version complète avec recherche multi-catégories
3. **Selenium\_Complet.py**: Interface graphique principale

### 4- Architecture Modulaire

#### 1. Interface Graphique (AppGUI) - **Selenium\_Complet.py**

- **Technologies**: Tkinter pour l'interface, PIL pour la gestion des images
- **Fonctionnalités**:
  - Interface moderne avec animations
  - Trois types de recherche (multi-catégories, par catégorie, nouvelles catégories)
  - Affichage des résultats dans un tableau interactif
  - Fonctions d'export (Word, Excel, CSV, JSON, carte interactive)
  - Visualisation des détails des centres sélectionnés
  - Organisation en 3 sections :
    - Barre de titre/logo
    - Panneau de recherche
    - Résultats
    - Options d'export.
  - Interaction avec deux moteurs de scraping :
    - WasteSearchApp (recherche complète multi-catégorie)

- CentreTriScraper (recherche simple ou avancée)

## 2. Module Principal (Selenium\_auto\_Complet)

- **Technologies:** Selenium pour le web scraping, SQLite pour le stockage local
- **Composants clés:**
  - WasteSearchApp: Gère la recherche complète sur les 11 catégories de déchets
  - WasteSearchManager: Effectue les recherches par type de déchet
    - Collecte des données XML après interaction Selenium.
    - Remplissage auto de l'adresse (géolocalisation).
    - Sélection dynamique des types de déchets via CSS selectors.
  - AddressManager: Gère la géolocalisation et la validation des adresses
  - WasteExporter: Gère l'export des données dans différents formats
  - CentreTriDB: Base de données SQLite pour stocker les résultats avec dédoublonnage automatique.

## 3. Module Simplifié (Selenium\_auto\_Simple)

- **Fonctionnalités:** Version simplifiée et plus interactive
  - Recherche par catégorie unique ou par type de plateforme / type de déchets
  - Interface simplifiée pour des recherches ciblées
  - L'utilisateur choisit manuellement le **type de centre à scraper** (centre, déchèterie, collecteur, plateforme).
  - Système de sélection dynamique et persistante via fenêtre Tkinter.
  - Reprise de la base SQLite (centres\_tri.db) pour stockage local des résultats.
  - Mêmes fonctionnalités d'export que le module complet

## 5- Points Forts du Projet ([cliquez ici](#))

- ★ **Robustesse:** Gestion des erreurs et reprises automatiques
- ★ **Modularité:** Séparation claire des responsabilités
- ★ **Extensibilité:** Facile à modifier pour ajouter de nouvelles fonctionnalités
- ★ **Convivialité:** Interface utilisateur intuitive avec feedback visuel

### 1. Interface Utilisateur Complète:

- Design moderne avec animations fluides
- Organisation claire des fonctionnalités
- Visualisation interactive des résultats

### 2. Fonctionnalités Avancées de Scraping:

- Gestion des cookies et des en-têtes HTTP
- Récupération des données via XML
- Correction automatique des problèmes d'encodage

### **3. Système de Stockage:**

- Base de données SQLite intégrée
- Prévention des doublons
- Historique des recherches

### **4. Options d'Export Polyvalentes:**

- Formats variés (Word, Excel, CSV, JSON)
- Carte interactive avec Folium
- Mise en page professionnelle dans les exports

### **5. Gestion des Erreurs:**

- Messages d'erreur explicites
- Tentatives de récupération automatique
- Fenêtres de chargement pendant les opérations longues

## **6- Axes d'Amélioration Possibles**

### **1. Performance:**

- Optimiser les temps de chargement
- Implémenter un système de cache pour les recherches répétées

### **2. Fonctionnalités:**

- Ajouter des filtres avancés sur les résultats
- Ajouter des statistiques sur les résultats

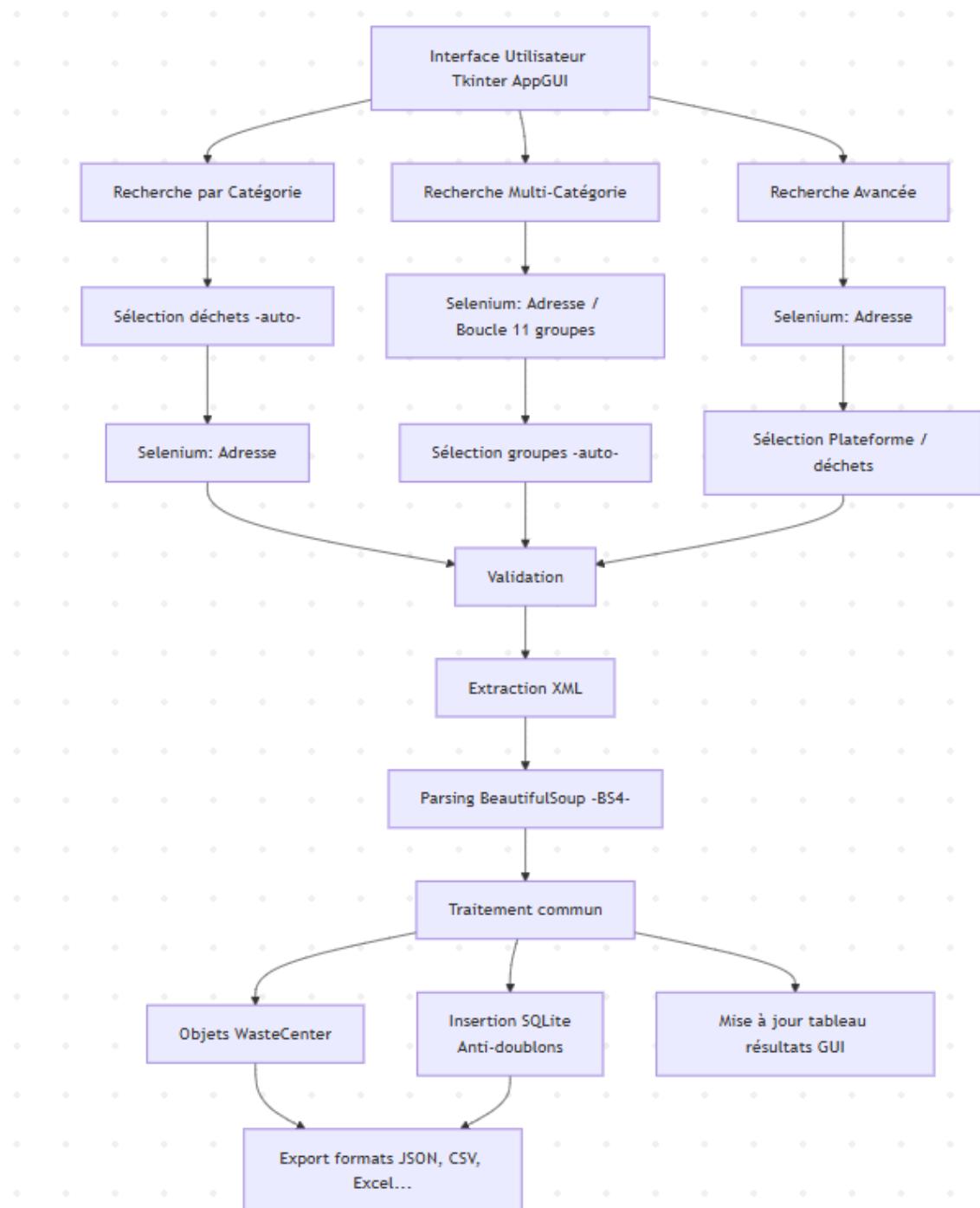
### **3. Maintenance:**

- Documentation plus complète
- Tests automatisés pour vérifier la stabilité

## 7- Fonctionnalités techniques principales

Fonction	Description
<b>Base SQLite</b>	Stockage des résultats avec historique + vérification de doublons.
<b>Exports</b>	JSON, CSV, Word (Docx), Excel, Carte (Folium).
<b>Animations UI</b>	Ouverture progressive des fenêtres pour une meilleure UX.
<b>Multithreading</b>	Threads utilisés pour ne pas bloquer l'UI pendant les longues requêtes.
<b>Parsing XML</b>	Utilisation de requests + BeautifulSoup pour traiter les réponses XML.
<b>Scraping Selenium</b>	Interaction directe avec le site FFB (formulaire ASP.NET complexe).
<b>Sélection dynamique</b>	Choix interactif du type de centre à scraper lors des recherches avancées.

## 8- Schéma fonctionnel – Flux global du projet ([cliquez ici](#))



## 9- Technologies Principales

Catégorie	Technologies
<b>Langages</b>	Python (pour l'ensemble du projet)
<b>Interface GUI</b>	Tkinter (bibliothèque standard Python pour l'interface graphique)
<b>Web Scraping</b>	Selenium WebDriver (automatisation navigateur)
<b>Parsing XML</b>	BeautifulSoup (analyse des données XML/HTML)
<b>Base de Données</b>	SQLite (stockage local des résultats)

## 10- Bibliothèques Python Clés

Bibliothèque	Utilisation
<b>requests</b>	Envoi de requêtes HTTP pour récupérer les données XML
<b>selenium</b>	Interaction avec les pages web (remplissage de formulaires, clics, etc.)
<b>bs4 (BeautifulSoup)</b>	Extraction et parsing des données XML/HTML
<b>sqlite3</b>	Gestion de la base de données SQLite

<b>pandas</b>	Export des résultats vers CSV/Excel (si utilisé pour l'export)
<b>json</b>	Export des résultats au format JSON
<b>python-docx</b>	Génération de rapports Word (si export DOCX implémenté)
<b>folium/geopy</b>	Génération de cartes interactives (si fonctionnalité cartographique présente)

## 11- Outils et Protocoles

Outil/Protocole	Rôle
<b>Performance API</b>	Récupération des URLs XML après soumission des formulaires
<b>Cookies/Headers</b>	Gestion des sessions HTTP pour contourner les protections anti-scraping
<b>WebDriver</b>	Pilotes pour Selenium (ChromeDriver, GeckoDriver, etc.)

## 12- Formats d'Export Supportés

- Fichiers : JSON, CSV, Excel (via `pandas`), Word (via `python-docx`)
- Cartographie : Bibliothèques comme `folium` pour des cartes Leaflet (si implémenté)
- Base de données : SQLite pour un stockage persistant local

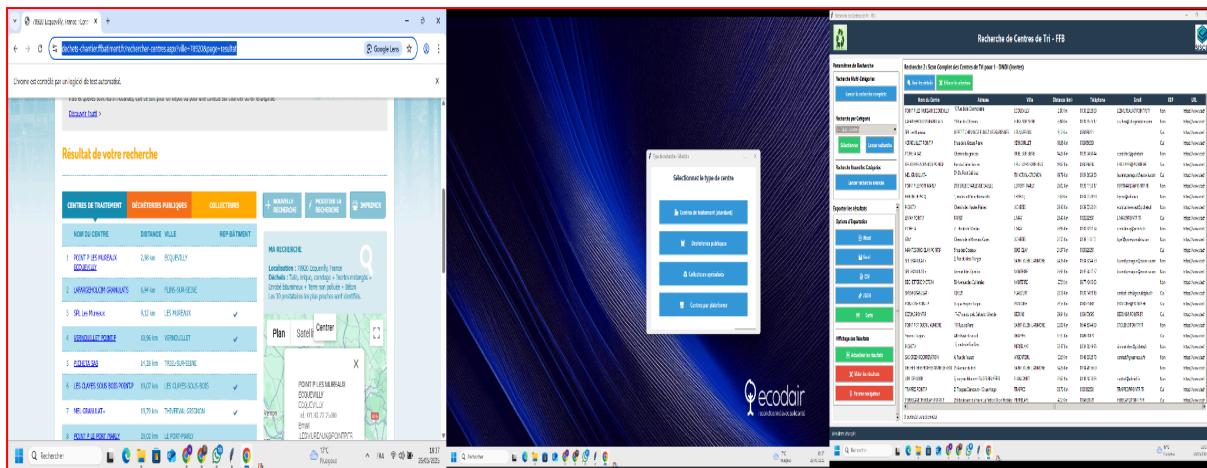
## 13 - Schéma Technologique Résumé ([cliquez ici](#))

Python Core

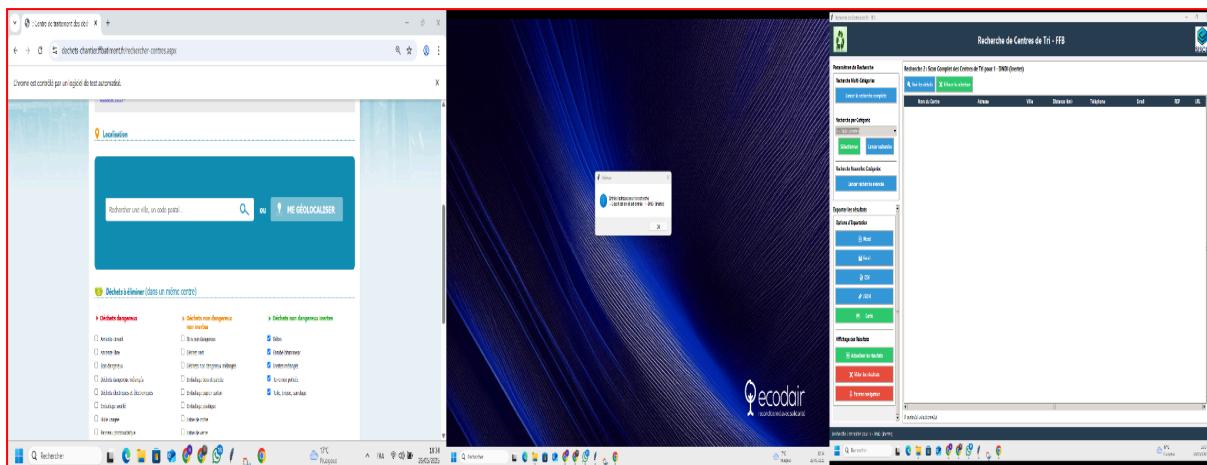
- └── Tkinter (GUI)
- └── Selenium (Scraping)
  - └── WebDriver (Chrome/Firefox)
  - └── Performance API (Capture XML)
- └── BeautifulSoup (Parsing)
- └── SQLite (DB)
- └── Export Modules
  - └── pandas (CSV/Excel)
  - └── python-docx (Word)
  - └── folium (Cartes)



L'interface principale du application Scraping



## Résultat de recherche



## Entrer L'adresse



#### Affichage le nombre totale de recherche complet

Détails des Centres de Tri

## Centre 1

### LAFARGEHOLCIM GRANULATS

#### Informations générales

Adresse: 11 Rue des Chevries  
 Ville: FLINS-SUR-SEINE  
 Distance: 6,94 Km km  
 Téléphone: 01 30 95 75 12  
 Email: css.flins@lafargeholcim.com

#### Caractéristiques

Partenaire REP: Non  
 Type de centre: Centre de traitement  
 Déchets acceptés: Tuile, brique, carrelage + Inertes mélangés + Enrobé bitumineux + Terre non polluée + Béton  
 Coordonnées: 48.97199592356605, 1.859333443130481

[Ouvrir la page web](#) [Copier les coordonnées](#) [Afficher sur carte](#)

[Fermer](#)

### Affichage Détailé du centre

carte maps.html

Centres de traitement des déchets par catégorie

- 1 - DNDI (inertes)
- 2 - Métaux (DNDNI)
- 3 - Bois (DNDNI)
- 4 - Menusries vitrées (DNDNI)
- 5 - Plâtre (DNDNI)
- 6 - Plastique (DNDNI)
- 7 - Laine de verre/ de roche/ polystyrène (DNDNI)
- 8 - Autres DNDNI
- 9 - Amiante ciment (DD)
- 10 - DEEE
- 11 - Autres Matériaux dangereux (DD)

### Génération de cartes géographique

**Résultats des recherches de centres de traitement**

Date : 26/05/2025 18:46

**1 - DNDI (inertes) :**

Localisation : 7890 Av. des Pyrénées, 33114 Le Barp, France.  
Déchets : Tuile, brique, carrelage + Inertes mélangés + Terre non polluée + Béton

Nom	Adresse	Ville	Distance	Téléphone	Email	REP Batiment	Lien
PENA ENVIRONNEMENT	4773 avenue de Pierrotin	SAINT-JEAN-D'ILLAC	14,46 Km	05 57 97 37 77	pena@wanadoo.fr	Non	PENA ENVIRONNEMENT
Ecoreval	ZI de croix d'hins	MARCHEPRI ME	17,51 Km	05 56 17 40 73	contact@ecopool.fr	Non	Ecoreval
SUEZ RV Sud Ouest BTP	321, allée de Péronette	SAINT-JEAN-D'ILLAC	22,39 Km	05 56 21 36 42	st.leroy@suez.com	Non	SUEZ RV Sud Ouest BTP
XEROS ENVIRONNEMENT	134 allée COURBET	SAINT-JEAN-D'ILLAC	23,79 Km	06 71 23 08 42	contact@xeros-environnement.fr	Non	XEROS ENVIRONNEMENT
PENA ENVIRONNEMENT	Chemin de la Poudrière	MERIGNAC	25,39 Km	05 57 97 37 77	pena@wanadoo.fr	Non	PENA ENVIRONNEMENT
Guyenne Environnement	Passe des villas	MERIGNAC	26,6 Km	05 56 13 11 57	bascule.meriguenne-environnement.com	Non	Guyenne Environnement
AUDOIN et Fils Bétons	Passage Luchbert	MERIGNAC	27,12 Km	05 56 34 16 28	N/A	Non	AUDOIN et Fils Bétons

### Export des Résultats : sous forme Word

**fch1.xlsx — LibreOffice Calc**

Fichier Édition Affichage Insertion Format Styles Feuille Données Outils Fenêtre Aide

Calibri 11 pt G I S A

A2 7890 Av. des Pyrénées, 33114 Le Barp, France

A	B	C	D	E	F
Localisation	nom_centre	adresse_centre	ville_centre	distance_centre	tel_centre
7890 Av. des Pyrénées, 33114 Le Barp, France	PENA ENVIRONNEMENT	4773 avenue de Pierrotin ZI de croix d'hins Avenue de la Mission Blanche	SAINT-JEAN-D'ILLAC	14,46	05 57 97 37 77
7890 Av. des Pyrénées, 33114 Le Barp, France	Ecoreval		MARCHEPRI ME	17,51	05 56 17 40 73
7890 Av. des Pyrénées, 33114 Le Barp, France	SUEZ RV Sud Ouest BTP	321, allée de Péronette	SAINT-JEAN-D'ILLAC	22,39	05 56 21 36 42
7890 Av. des Pyrénées, 33114 Le Barp, France	XEROS ENVIRONNEMENT	134 allée COURBET	SAINT-JEAN-D'ILLAC	23,79	06 71 23 08 42
7890 Av. des Pyrénées, 33114 Le Barp, France	PENA ENVIRONNEMENT	Chemin de la Poudrière	MERIGNAC	25,39	05 57 97 37 77
7890 Av. des Pyrénées, 33114 Le Barp, France	Guyenne Environnement	Passe des villas	MERIGNAC	26,6	05 56 13 11 57
7890 Av. des Pyrénées, 33114 Le Barp, France	AUDOIN et Fils Bétons	Passage Luchbert	MERIGNAC	27,12	05 56 34 16 28
7890 Av. des Pyrénées, 33114 Le Barp, France	VEOLIA BEGLES	11 Rue Louis Bérriot, 33130 Begles	BEGLES	28,21	05 56 49 75 00
7890 Av. des Pyrénées, 33114 Le Barp, France	Ecobenne	123 avenue de Capéyron	SAINT-MEDARD-EN-JALLES	33,21	05 56 17 40 73
7890 Av. des Pyrénées, 33114 Le Barp, France	SOCIETE NOUVELLE CHALLENGER	Route de Blagnon	LANTON	37,52	0556828150
7890 Av. des Pyrénées, 33114 Le Barp, France	TRANSPORTS BAQUEY	3, avenue de Braude	LE TAILLAN-MEDOC	37,89	0556354441
7890 Av. des Pyrénées, 33114 Le Barp, France	Ecotri Andernos	28 avenue Gustave Eiffel	ANDERNOS-LES-BAINS	43,88	05 56 17 40 73
7890 Av. des Pyrénées, 33114 Le Barp, France	SUEZ/VAL+	ZI La chataigneraie Ave Lassalle de Ciron	PREIGNAC	43,91	05 56 76 29 21
7890 Av. des Pyrénées, 33114 Le Barp, France	SDC AMBARES	Avenue des industries	AMBARES-ET-LAGRAVE	48,03	0620970970
7890 Av. des Pyrénées, 33114 Le Barp, France	AZURA RECYCLAGE - Site de Bassens	ZI DES GUERLANDES, AV DES GUERLANDES	BASSENS	48,43	0556385444
7890 Av. des Pyrénées, 33114 Le Barp, France	LES PIERRES DE FRONTENAC		JUGAZAN	65,16	055755832
7890 Av. des Pyrénées, 33114 Le Barp, France	SEGM	La Bordette	MERLES	70	05 63 94 81 86
7890 Av. des Pyrénées, 33114 Le Barp, France	BCCL	RD145 LD Le Terrier de Pierre Folle	BEDENAC	80,46	05 46 04 47 84
7890 Av. des Pyrénées, 33114 Le Barp, France	INERTAM-COFAL	471, route de Cantegrit Est	MORCENX	86,03	05 58 04 17 49
7890 Av. des Pyrénées, 33114 Le Barp, France	TGF - Tri Garonne Environnement	1705 Porte de Gironde	SAINTE-BAZEILLE	86,13	0553848138
7890 Av. des Pyrénées, 33114 Le Barp, France	VEOLIA PROPRETE	345, route Jean d'Arnaud	LALIQUE	108,32	05 58 57 98 98
7890 Av. des Pyrénées, 33114 Le Barp, France	CLTDI	Route de la Lande	BESGAAR	108,91	05 58 43 92 91
7890 Av. des Pyrénées, 33114 Le Barp, France	Société départementale de carrières	1358 chemin de Bourrèque	BUZET-SUR-BAISE	111,4	05 53 79 46 01
7890 Av. des Pyrénées, 33114 Le Barp, France	BCCL	Avenue de Plaisance	BARBEZIEUX-SAINT-HILAIRE	113,56	05 45 98 34 54
7890 Av. des Pyrénées, 33114 Le Barp, France	BCCL	53 route de Saint Genis	SAINTE-GERMAINE-DE-LUSIGNAN	114,38	05 46 48 15 45
7890 Av. des Pyrénées, 33114 Le Barp, France	DECHETERIE DE SIGOULES		SIGOULES	126,02	0553233228
7890 Av. des Pyrénées, 33114 Le Barp, France	CLTDI	300, Rue Monge	SAINTE-AVIT	133,2	05 58 85 29 74
7890 Av. des Pyrénées, 33114 Le Barp, France	DECHETERIE DE RABIE		SAINTE-LIVRADE-SUR-LOT	138,65	0553010204

### Export des Résultats : sous forme Excel

```

fch1.json
{
  "metadata": {
    "generated": "2020-05-26T18:47:29.373Z",
    "total_recherches": 11,
    "total_centres": 322
  },
  "results": [
    {
      "categorie": "1 - DNDI (Inertes)",
      "type_recherche": "Déchets : Tuile, brique, carrelage + Inertes mélangés + Enrobé bitumineux + Terre non polluée + Béton;7890 Av. des Pyrénées, 33114 Le Barp, France;PENA ENVIRONNEMENT;4773 avenue de Pierrotin;SAINT-JEAN-D'ILLAC;14,46;05 57 37 77;peña@wanadoo.fr;Non;44.769090;-0.769395
      ...
    }
  ]
}

fch1.csv
Catégorie;Type recherche;Localisation;Nom centre;Adresse;Ville;Distance (km);Téléphone;Email;REP Batiment;Latitude;Longitude
1 - DNDI (inertes);Déchets : Tuile, brique, carrelage + Inertes mélangés + Enrobé bitumineux + Terre non polluée + Béton;7890 Av. des Pyrénées, 33114 Le Barp, France;PENA ENVIRONNEMENT;4773 avenue de Pierrotin;SAINT-JEAN-D'ILLAC;14,46;05 57 37 77;peña@wanadoo.fr;Non;44.769090;-0.769395
1 - DNDI (inertes);Déchets : Tuile, brique, carrelage + Inertes mélangés + Enrobé bitumineux + Terre non polluée + Béton;7890 Av. des Pyrénées, 33114 Le Barp, France;Ecoreval;"ZI de croix d'hins Avenue de la Maison Blanche";MARCHEPRIME;17,51;05 56 17 40
...

```

### Export des Résultats : sous forme CSV, JSON

## Conclusion

Ce projet est une solution robuste et bien conçue pour la collecte et l'exploitation des données des centres de tri de la FFB. Son architecture modulaire et ses fonctionnalités complètes en font un outil précieux pour les professionnels du bâtiment. Il répond parfaitement à l'objectif initial de fournir un moyen efficace d'accéder et d'exploiter ces informations. Les améliorations potentielles concernent principalement l'optimisation et l'enrichissement des fonctionnalités existantes plutôt que des corrections fondamentales.