

Spécifications site web relatif aux espèces en danger - association OKFN France

Le présent document reprend les principaux éléments de conception de l'API relative aux espèces en danger. Il présente les principaux choix techniques, ceux relatifs à l'obtention des données externes ainsi que le modèle de données (schéma physique de données) de la base de données interrogée par l'API.

Benoît Prieur - 30 octobre 2020 - CC-BY-SA - V 0.1

1. La construction de la base de données

1.1 Aspects techniques

Le système de gestion de base de données utilisé est MySQL.

Le code Python de l'import des données depuis Wikidata, CITES et d'autres sources à venir est disponible à l'adresse suivante sur Github.

<https://github.com/okfnfr/CITES/tree/master/IMPORT%20DB>

Le code SQL (inclus dans du Python) relatif à la description des tables de la base de données est dans le fichier tables.py.

<https://github.com/okfnfr/CITES/blob/master/IMPORT%20DB/tables.py>

1.1 Fonctionnement général

- On commence par obtenir l'ensemble des données relatifs aux entrées de Wikidata possédant la propriété "identifiant Species+" (P2040) de Wikidata.
 - La propriété est à cette adresse <https://www.wikidata.org/wiki/Property:P2040>
 - La requête SPARQL est ici :
https://github.com/okfnfr/CITES/blob/master/IMPORT%20DB/C_WIKIDATA.py
 - Une fois cette opération réalisée, on a environ 25 000 espèces et nombre de leurs propriétés, y compris les identifiants de chaque espèce dans plusieurs bases externes, à commencer évidemment par CITES.
 - On alimente la table principale MAINTABLE avec ses données.

- Grâce à l'id Species+ (CITES) on requête via leur API dédiée des données spécifiques à CITES.
 - Pour le moment (au 30 octobre 2020), on requête uniquement les données relatives à la distribution et on les range dans la table CITES_DISTRIBUTION.

De manière générale, on aura pour enrichir la base de données, la procédure suivante :

- Wikidata nous "fournit" l'id d'une base de données externe pour une espèce donnée.
- Grâce à cet id, on requête cette base de données externe, on récupère les données et on les consolide dans une table prévue à cet effet.
- Ainsi l'enrichissement de la BDD et donc de l'API OKFNFR se fait par ajout de tables dédiées. Pas nécessairement une table par base de données externe d'ailleurs : ainsi on a pour CITES une table pour la distribution, puis on aura sans doute une autre table pour les aspects juridiques etc.

2. L'API

Une fois créé et alimenté, la base de données MySQL est mise en ligne. Elle est interrogeable par une API développée elle aussi en Python. L'ensemble base de données + API est actuellement hébergé sur Héroku.

Le code de l'API est à cette adresse :

<https://github.com/okfnfr/CITES/tree/master/API>

Dans l'hypothèse de l'intégration d'une nouvelle source de données externe, on procède ainsi :

1. Wikidata nous fournit l'id dans cette source de données. Cet id est stocké dans MAINTABLE.
2. Une table est créée dans le modèle de données, relative à l'apport de cette source de données (autres aspects de répartition, images, sons etc.).
3. Les points d'entrée permettant de spécifiquement interroger ces aspects sur l'API doivent être ajoutés (point d'entrée + requêtes SQL correspondantes).

Au 30 octobre 2020, les requêtes HTTP GET disponibles sont les suivantes :

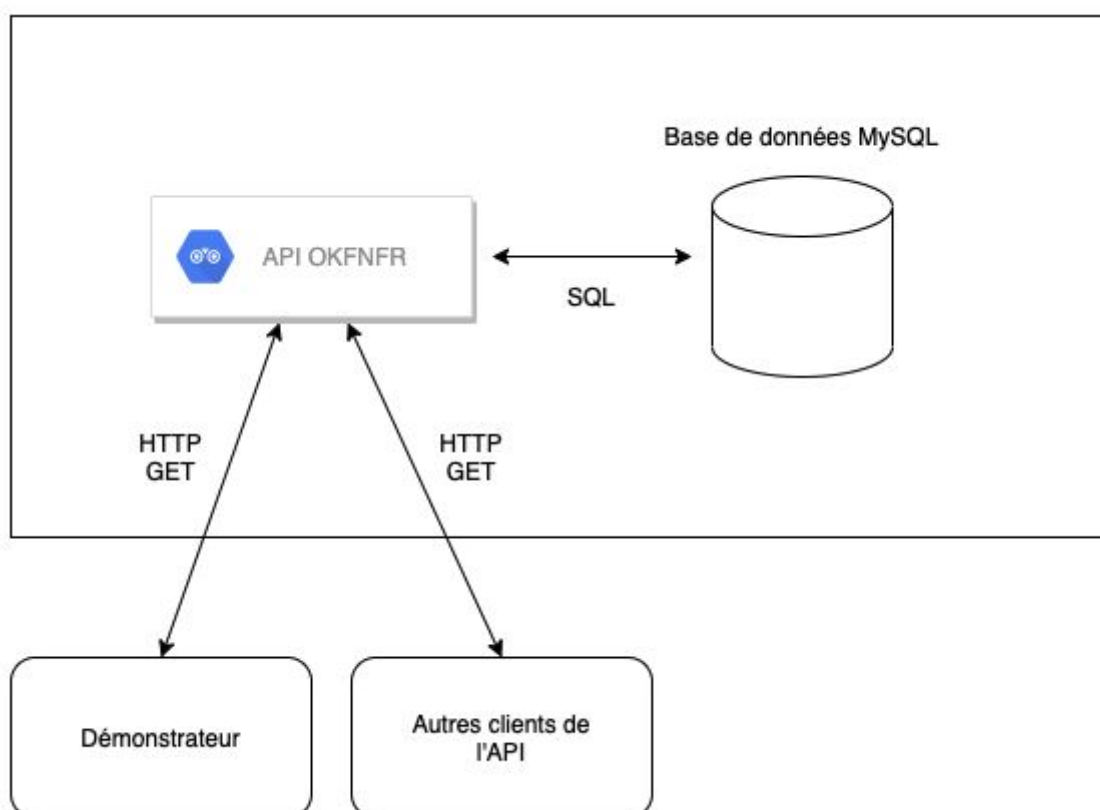
1. */countries/ : obtenir la liste ISO des pays*
2. */country/<string:iso> : obtenir un pays par son code ISO*
3. */itemWikidata/<string:wikidataid> : obtenir une espèce depuis son identifiant Wikidata*
4. */itemCites/<string:citesid> : obtenir une espèce depuis son identifiant CITES*
5. */SearchTextBegin/<string:beginText> : obtenir une espèce dont le nom scientifique ou non commence par...*
6. */SearchIncludeText/<string:includetext> : obtenir une espèce dont le nom scientifique ou non inclut la chaîne de caractères...*

7. /ByCountryCode/<string:countrycode> : *obtenir les espèces menacés CITES dans un pays donné, par son code ISO*
8. /ByCountryName/<string:countryname> : *obtenir les espèces menacés CITES dans un pays donné, par son nom*

Pour 2 et 3, la liste des pays au sens de CITES est remontée dans les résultats.

3. Schéma global et modèle de données

3.1 Schéma global



3.2 Modèle de données (au 30 octobre 2020)

