



ÉCOLOGIE &
ENVIRONNEMENT

Comparer et partager ses données

Chloé MARTIN - UAR 2047 DoHNEE Données de recherche pour
l'Histoire Naturelle, l'Écologie et l'Environnement
Stefan GAGET – UMR 8199 EGID European Genomic Institute for
Diabetes



Sommaire

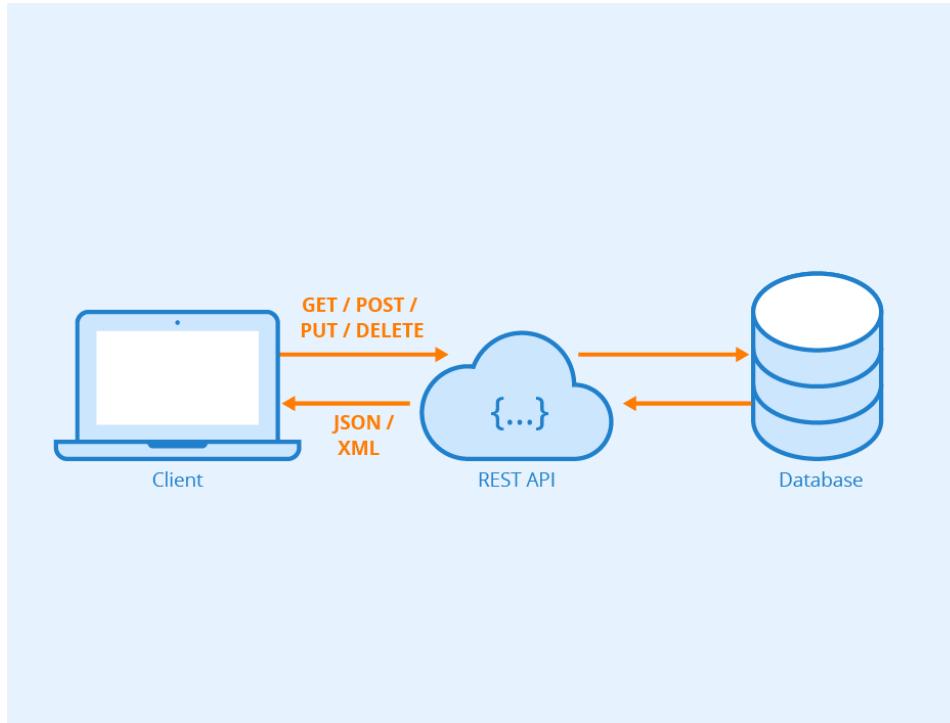
Enrichir ses données avec une API

Aligner avec un autre projet

Réconcilier ses données avec Wikidata

Enrichir ses données avec une API

API = interface de programmation d'application



API GBIF

- doc générale : <https://techdocs.gbif.org/en/openapi/>
- doc espece : <https://techdocs.gbif.org/en/openapi/v1/species>
- URL permettant d'accéder aux informations d'espèces :
- <http://api.gbif.org/v2/species/match?verbose=true&name=espece>

API GBIF

```
value.parseJson().get("kingdom")
+",
"+value.parseJson().get("phylum")
+",
"+value.parseJson().get("class")+
",
"+value.parseJson().get("order")+
, "+value.parseJson().get("family")
```

JSON Données brutes En-têtes

Enregistrer Copier Tout réduire Tout développer ▾ Filtrer le JSON

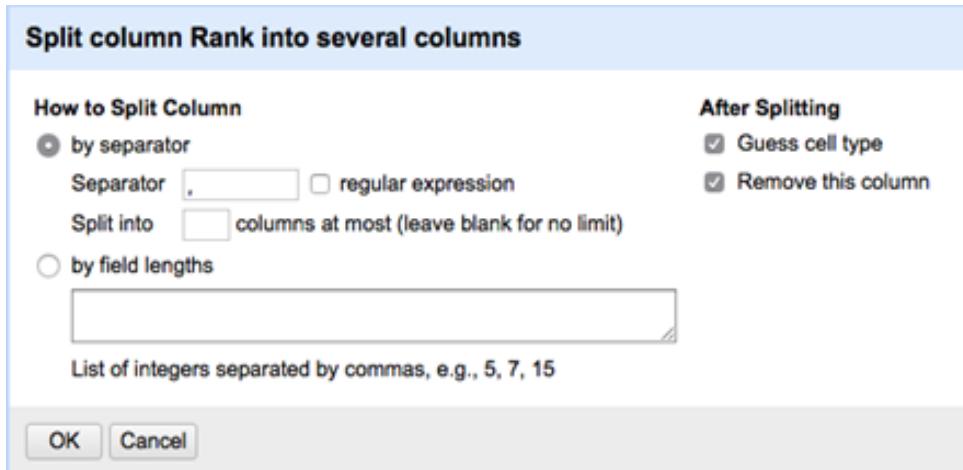
```
usageKey: 5219142
scientificName: "Canis Linnaeus, 1758"
canonicalName: "Canis"
rank: "GENUS"
status: "ACCEPTED"
confidence: 93
note: "Similarity: name=100; authorship=0; kingdom=0; rank=0; status=1; score=101; nextMatch=0"
matchType: "EXACT"
alternatives:
  ▾ 0:
    usageKey: 8227762
    acceptedUsageKey: 2434453
    scientificName: "Canis Sclater, 1883"
    canonicalName: "Canis"
    rank: "GENUS"
    status: "SYNONYM"
    confidence: 93
    note: "Similarity: name=100; authorship=0; kingdom=0; rank=0; status=0; score=100"
    matchType: "EXACT"
    kingdom: "Animalia"
    phylum: "Chordata"
    order: "Carnivora"
    family: "Canidae"
    genus: "Canis"
  kingdomKey: 1
  phylumKey: 44
  classKey: 359
  orderKey: 732
  familyKey: 9701
  genusKey: 2434453
  synonym: true
  class: "Mammalia"
kingdom: "Animalia"
phylum: "Chordata"
order: "Carnivora"
family: "Canidae"
genus: "Canis"
kingdomKey: 1
phylumKey: 44
```

TP :

- **1**-Aller sur Collector, puis effectuer un Text facet.
Sélectionner le collecteur Elsa P
- **2**-Sous Full name, cliquer sur le menu de la colonne Edit column > Add column by fetching URLs..., nommer la nouvelle colonne Api_name
- **3-Coller l'expression suivante:**
"[http://api.gbif.org/v1/species/match?verbose=true&name=+escape\(value,'url'\)](http://api.gbif.org/v1/species/match?verbose=true&name=+escape(value,'url'))"

TP :

- **4-Aller** sur Api_name, cliquer sur le menu de la colonne, puis sur Edit column > Add column based on this column.... Nommer la nouvelle colonne Rank et coller les expressions suivantes:
- value.parseJson().get("kingdom")+", "+value.parseJson().get("phylum")+", "+value.parseJson().get("class")+", "+value.parseJson().get("order")+", "+value.parseJson().get("family")
- **5-Sous Rank** suivre le menu Edit column > Split into several columns...



Aligner avec un autre projet

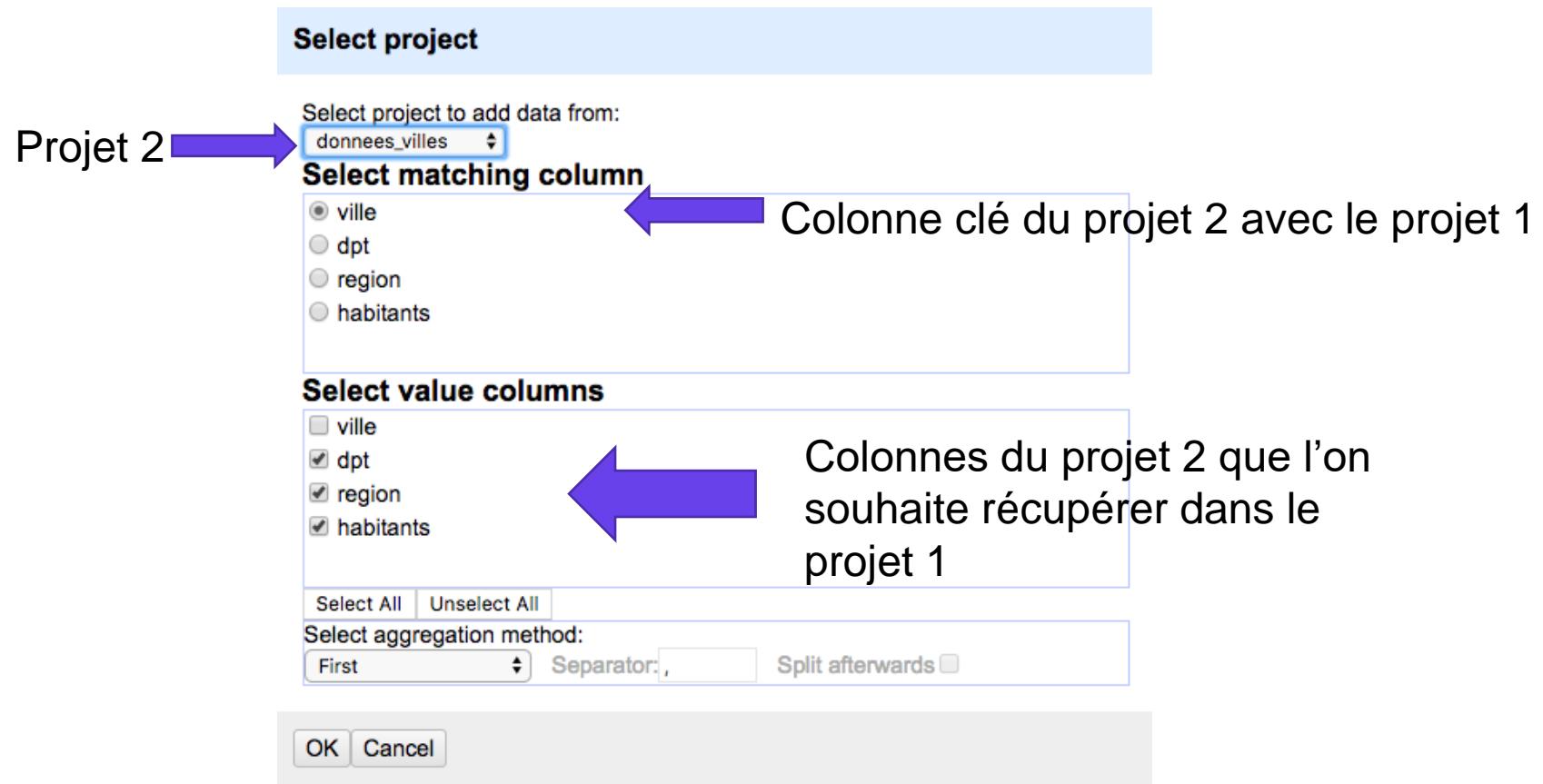
Utilisation du plugin VIB-Bits

- **Lien vers le plugin :**
<http://web.archive.org/web/20210817193059/http://data.bits.vib.be/hidden/g7dt6RjuUTU421dY2CwrGePGX/vib-bits.zip>
- **Démo en vidéo :** <https://www.canal-u.tv/chaines/rbdd/installer-une-extension-et-vib-bits>

Objectif : A partir de la colonne ville d'un projet 1, aligner avec la colonne ville du projet 2 et obtenir les informations associées contenues dans le projet 2

Toutes	code_personne	date	ville
1.	P001	01/02/2017	Nice
2.	P002	01/03/2017	Caen
3.	P003	15/02/2017	Lyon
4.	P004	15-02-2017	Nice
5.	P005	15-04-2017	Le Havre
6.	P005	12-02-2017	Le Havre
7.	P005	11/01/2017	
8.	P006	19/02 (2017)	Lyon
9.	P002	16/03 (2017)	Caen
10.	P005	08/01/2017	Le Havre

Aligner avec un autre projet



Aligner avec un autre projet

On obtient

Projet 1 Informations importées du projet 2

ville	dpt	region	habitants
Nice	Alpes-Maritimes	Provence-Alpes-Côte d'Azur	342000
Caen	Calvados	Normandie	106000
Lyon	Rhône	Auvergne-Rhône-Alpes	513000
Nice	Alpes-Maritimes	Provence-Alpes-Côte d'Azur	342000
Le Havre	Seine-Maritime	Normandie	172000
Le Havre	Seine-Maritime	Normandie	172000
Lyon	Rhône	Auvergne-Rhône-Alpes	513000
Caen	Calvados	Normandie	106000
Le Havre	Seine-Maritime	Normandie	172000

INSTALLATION VIB_BITS

Lien vers les extensions :

<https://openrefine.org/extensions>

The screenshot shows the OpenRefine interface version 3.9.2 [TRUNK]. The left sidebar has a red box around the 'Ouvrir un projet' (Open project) button. The main area displays a list of projects in a table with columns: Nom (Name), Modifié le (Modified on), and Typ (Type). A large blue arrow points from the sidebar towards the table. Another blue arrow points from the 'vib-bits' folder entry in the table towards the right. A red box highlights the 'vib-bits' folder entry in the table. The table data includes:

Nom	Modifié le	Typ
2015 2	07/10/2024 10:41	Dossier de fichiers
AMPC		Dossier de fichiers
AMPC		Dossier de fichiers
feu zoi		Fichier JSON
metadat		Fichier JSON
FAST		Fichier JSON
fast GS		Fichier JSON
fast GS		Fichier JSON

+ Redémarrer l'application !

Réconcilier ses données

Qu'est ce que la réconciliation ?

<https://www.wikidata.org/wiki/Q535>

Victor Hugo (Q535)



Référentiel auteur de la BNF
VIAF
etc

<https://www.wikidata.org/wiki/Q142>

France (Q142)



Codes de pays définis par l'ISO
Vocabulaire Rameau
etc

Réconcilier ses données

Accéder au service :

17 lignes

Voir en: lignes entrées Afficher: 5 10 25 50 lignes

			nom_scientifique	nom_vernacula	famille
1.	67778	Facette	Truite de rivière	Salmonidae	
2.	67854	Filtrer le texte	Corégone blanc	Salmonidae	
3.	67812	Éditer les cellules	Omble chevalier	Salmonidae	
4.	67804	Éditer la colonne	Truite arc-en-ciel	Salmonidae	
5.	67819	Transposer	Cristivomer	Salmonidae	
6.	67759	Trier...	Ombre commun	Salmonidae	
7.	67794	Aperçu	Luchon	Salmonidae	
8.	67765		Saumon atlantique	Salmonidae	
9.	67111		Aspilette	Cyprinidae	
10.	67466	Réconcilier	Démarrer la réconciliation...		

Extract named entities... Facettes

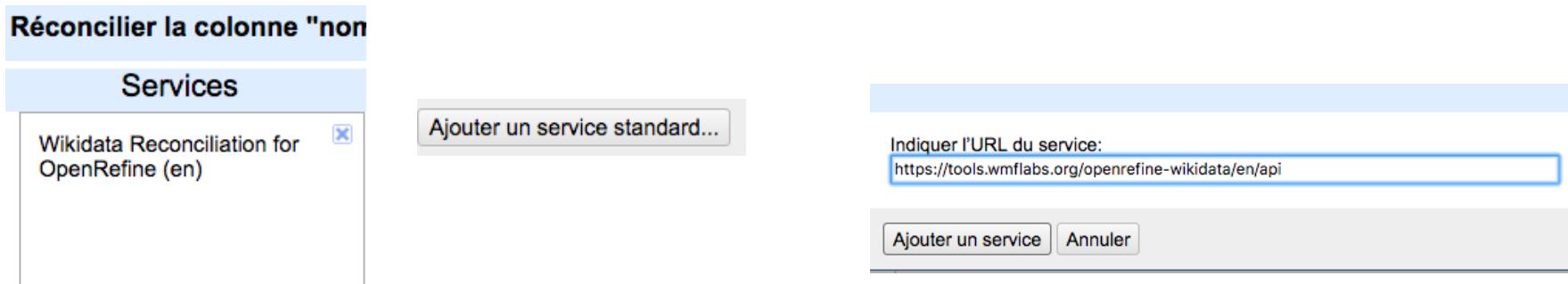
Actions

Copier les données de reconciliation...

Réconcilier ses données

Adresse de l'API wikidata

**<https://tools.wmflabs.org/openrefine-wikidata/en/api> ou
[/fr/api](https://tools.wmflabs.org/openrefine-wikidata/fr/api)**



Réconcilier ses données

Sélectionner le type d'entité **commun aux éléments de chaque ligne**.
Openrefine analyse les données et propose plusieurs choix. Si un choix n'est pas satisfaisant vous pouvez choisir manuellement un type

Réconcilier chaque cellule avec une entité de l'un de ces types:

- human
Q5
- book
Q571
- metro station
Q928830
- avenue
Q7543083
- film
Q11424

Utiliser également les détails pertinents des autres colonnes:

Colonne Inclure? Comme propriété
metier

Réconcilier avec le type: poet poet

Réconcilier sans type particulier

Correspondance automatique des valeurs candidates Auto-match candidates with high confidence

Nombre maximal de candidats renvoyés

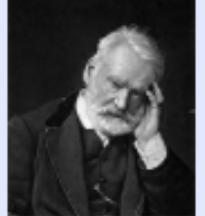
[Ajouter un service standard...](#) Démarrer la réconciliation Annuler

Réconcilier ses données

nom	nom-reconcilie	metier
Emile Ajar	Romain Gary Choisir une nouvelle correspondance	novelist
victor hugo	<input checked="" type="checkbox"/> Victor Hugo López Romero (100) <input checked="" type="checkbox"/> Victor Hugo Peña (81) <input checked="" type="checkbox"/> Victor Hugo Green (79) <input checked="" type="checkbox"/> Victor Hugo Cejas (79) <input checked="" type="checkbox"/> Victor Hugo Islas (79) <input checked="" type="checkbox"/> Victor Hugo Antelo (76) <input checked="" type="checkbox"/> Victor Hugo Ávalos (76) <input checked="" type="checkbox"/> Víctor Hugo Monzón (76) <input checked="" type="checkbox"/> Víctor Hugo Vieyra (76) <input checked="" type="checkbox"/> Víctor Hugo Rivera (76) <input checked="" type="checkbox"/> Víctor Hugo Marco (76) <input checked="" type="checkbox"/> Víctor Hugo García (76) <input checked="" type="checkbox"/> Víctor Hugo Amatti (76) <input checked="" type="checkbox"/> Víctor Hugo Carrizo (73) <input checked="" type="checkbox"/> Víctor Montaño (73) <input checked="" type="checkbox"/> Hugo Benioff (73) <input checked="" type="checkbox"/> Victor Hugo Basinet (73) <input checked="" type="checkbox"/> Víctor Hugo Morales Pérez (73) <input checked="" type="checkbox"/> Victor Hugo Halperin (71) <input checked="" type="checkbox"/> Víctor Hugo Carrillo (71) <input checked="" type="checkbox"/> Víctor Hugo Cárdenas (71) <input checked="" type="checkbox"/> François-Victor Hugo (71) <input checked="" type="checkbox"/> Créer un nouveau sujet Chercher les correspondances	novelist

Equivalent d'une coche simple

Apparier cette cellule



Victor Hugo (Q535)

French poet, novelist, and dramatist

nom	nom-reconcilie	metier
Emile Ajar	Romain Gary Choisir une nouvelle correspondance	novelist
victor hugo	Victor Hugo Choisir une nouvelle correspondance	novelist

Réconcilier ses données

Pour récupérer des données à partir de wikiData

A partir de la version 2.8 : « Editer la colonne > ajouter des colonnes à partir de valeurs réconciliées » puis choisir une propriété

Ex : date of birth

Ex: child

Add Property Preview

<input type="text" value="date of birth"/>											
Select an item from the list:											
date of birth	P569										
<table border="1"> <thead> <tr> <th>▼ nom</th> <th>▼ date of birth</th> </tr> </thead> <tbody> <tr> <td>Romain Gary</td> <td>1914-05-21T00:00:00Z</td> </tr> <tr> <td>Choisir une nouvelle correspondance</td> <td></td> </tr> <tr> <td>Victor Hugo</td> <td>1802-02-26T01:00:00Z</td> </tr> <tr> <td>Choisir une nouvelle correspondance</td> <td></td> </tr> </tbody> </table>		▼ nom	▼ date of birth	Romain Gary	1914-05-21T00:00:00Z	Choisir une nouvelle correspondance		Victor Hugo	1802-02-26T01:00:00Z	Choisir une nouvelle correspondance	
▼ nom	▼ date of birth										
Romain Gary	1914-05-21T00:00:00Z										
Choisir une nouvelle correspondance											
Victor Hugo	1802-02-26T01:00:00Z										
Choisir une nouvelle correspondance											

child	
Select an item from the list:	
child	P40
nom	child supprimer configurer
Romain Gary	Alexandre Diego Gary
Victor Hugo	Adèle Hugo
	Charles Hugo
	François-Victor Hugo
	Léopoldine Hugo

TP 2 : Traiter les fichiers des données de suivis d'oiseaux aux TAAF

Dossier TP2

Objectifs

- Nettoyer les données pour les 10 fichiers
- Réconcilier la colonne espece
- Ajouter le nom d'espèce (et pas le nom vernaculaire)
- Ajouter Kingdom, Phylum, Class, Order et family pour chaque taxon à partir de l'API GBIF
- Ajouter l'id TAXREF
- Ajouter les idinds à partir du fichier catalogue métal
 - Optionnel faire la même chose à partir des bagues darvik (attention à prendre en compte la couleur et l'espèce car les numéros de darvik ne sont pas unique !)
- Exporter uniquement les nouvelles bagues métal en prenant soin d'exporter les informations nécessaires pour alimenter la catalogue des métal

Liste des noms d'espèces

- **Chionis**
- **Grand Albatros**
- **Océanite à croupion gris**
- **Océanite à ventre noir**
- **Océanite de Wilson**
- **Pétrel à menton blanc**
- **Pétrel à tête blanche (=Pétrel de lesson)**
- **Pétrel bleu**
- **Pétrel de Kerguelen**
- **Pétrel gris**
- **Pétrel noir**
- **Pétrel plongeur commun**
- **Pétrel plongeur de Géorgie du sud**
- **Prion de Belcher**
- **Skua subantarctique**

Licence

<https://creativecommons.org/licenses/by/4.0>



Attribution 4.0 International (CC BY 4.0)

Vous êtes autorisé à :

- **Partager** — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats
- **Adapter** — remixer, transformer et créer à partir du matériel pour toute utilisation, y compris commerciale.

Selon les conditions suivantes :

Attribution — Vous devez créditer l'Œuvre, intégrer un lien vers la licence et indiquer si des modifications ont été effectuées à l' Œuvre. Vous devez indiquer ces informations par tous les moyens raisonnables, sans toutefois suggérer que l'Offrant vous soutient ou soutient la façon dont vous avez utilisé son Œuvre.

Pas de restrictions complémentaires — Vous n'êtes pas autorisé à appliquer des conditions légales ou des [mesures techniques](#) qui restreindraient légalement autrui à utiliser l' Œuvre dans les conditions décrites par la licence.

Attribution : C. Martin & S. Gaget – Ecole thématique Data SEE-Life - 2025