



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

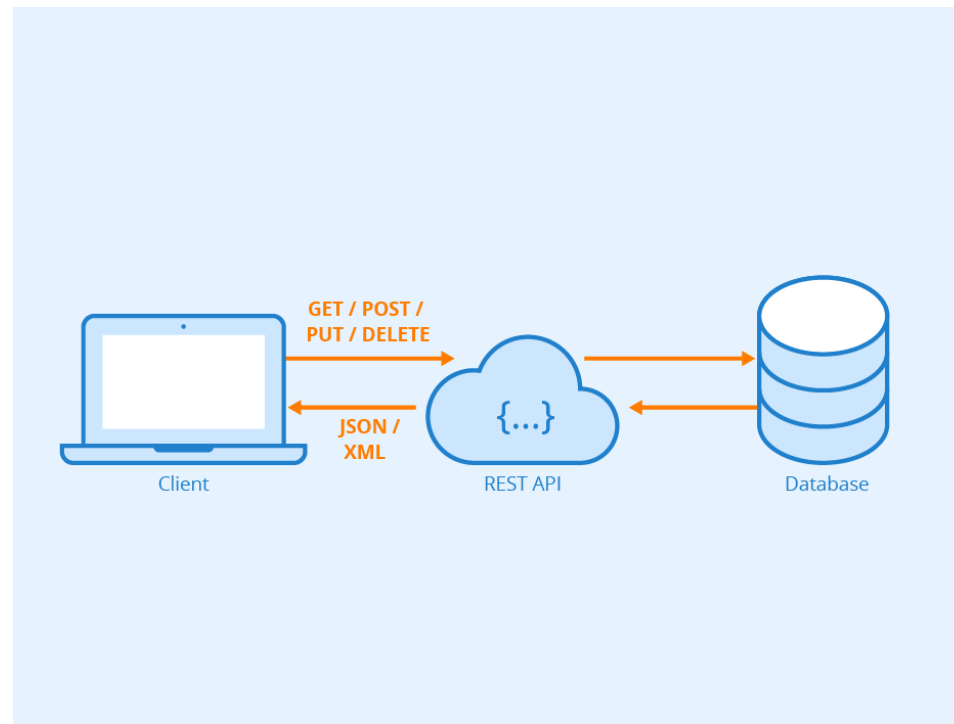
Enrichir ses données avec une API

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'especes :
- <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")
```

The screenshot displays a JSON viewer interface with tabs for 'JSON', 'Données brutes', and 'En-têtes'. Below the tabs are buttons for 'Enregistrer', 'Copier', 'Tout réduire', 'Tout développer', and a search filter 'Filtrer le JSON'. The JSON data is shown in a dark theme. The first entry (usageKey: 5219142) is for 'Canis Linnaeus, 1758' and is partially expanded. The second entry (usageKey: 8227762) is for 'Canis Sclater, 1883' and is also partially expanded. A third entry is shown below, which is boxed to highlight its taxonomic structure:

```
kingdom: "Animalia"
phylum: "Chordata"
order: "Carnivora"
family: "Canidae"
genus: "Atelocynus"
```

Below this, another set of keys is visible:

```
kingdomKey: 1
phylumKey: 44
classKey: 359
orderKey: 732
familyKey: 9701
genusKey: 2434453
synonym: true
class: "Mammalia"
```

At the bottom, more taxonomic data is shown:

```
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\(value,'url'\)"](#)

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...

Split column Rank into several columns

How to Split Column

☒ by separator
Separator ☐ regular expression
Split into columns at most (leave blank for no limit)

☐ by field lengths

List of integers separated by commas, e.g., 5, 7, 15

After Splitting

☒ Guess cell type
☒ Remove this column

OK Cancel

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

The screenshot shows the VIB-Bits interface. On the left is a table with columns: 'Toutes', 'code_personne', 'date', and 'ville'. The 'ville' column is highlighted with a red box. The table contains 10 rows of data. To the right of the table is a context menu with options: 'Facette', 'Filtrer le texte', 'Éditer les cellules', 'Éditer la colonne', 'Transposer', 'Trier...', 'Aperçu', and 'Réconcilier'. The 'Éditer la colonne' option is highlighted with a red box. To the right of the context menu is a sub-menu with options: 'Diviser en plusieurs colonnes...', 'Ajouter une colonne en fonction de cette colonne', 'Add column(s) from other projects...', and 'Ajouter une colonne en moissonnant des URL...'. The 'Add column(s) from other projects...' option is highlighted with a red box.

	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

Facette

Filtrer le texte

Éditer les cellules

Éditer la colonne

Transposer

Trier...

Aperçu

Réconcilier

Diviser en plusieurs colonnes...

Ajouter une colonne en fonction de cette colonne

Add column(s) from other projects...

Ajouter une colonne en moissonnant des URL...

Aligner avec un autre projet

Projet 2 →

Select project

Select project to add data from:
donnees_villes

Select matching column

☒ ville
☐ dpt
☐ region
☐ habitants

← Colonne clé du projet 2 avec le projet 1

Select value columns

☐ ville
☒ dpt
☒ region
☒ habitants

← Colonnes du projet 2 que l'on souhaite récupérer dans le projet 1

Select All Unselect All

Select aggregation method:
First Separator: , Split afterwards ☐

OK Cancel

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>

 **OpenRefine** *Un outil puissant pour travailler avec des*

Créer un projet
Ouvrir un projet
Importer un projet
Langues
Extensions


Version 3.9.2 [TRUNK]

Préférences
Aide
À propos

Parcourir le dossier de l'espace de travail

	Détails				
×	Détails	1 avr. 2025, 16:31:14	project	10/03/2025 09:03	Dossier de fichiers
×	Détails	1 avr. 2025, 15:24:11		07/10/2024 10:40	Dossier de fichiers
×				07/10/2024 10:41	Dossier de fichiers
×				02/04/2025 11:20	Fichier JSON
×	Détails	28 mars 2025, 14:50		02/04/2025 11:15	Fichier JSON
×	Détails	28 mars 2025, 14:11			
×	Détails	26 mars 2025, 14:21			
×	Détails	24 mars 2025, 13:41			
×	Détails	24 mars 2025, 11:01			
×	Détails	19 févr. 2025, 09:06:43	2015 2		
×	Détails	13 févr. 2025, 14:09:40	AMPC		
×	Détails	13 févr. 2025, 14:08:50	AMPC		
×	Détails	4 févr. 2025, 11:11:41	feu zo		
×	Détails	30 janv. 2025, 11:57:24	metad		
×	Détails	28 janv. 2025, 17:03:58	FAST		
×	Détails	28 janv. 2025, 16:35:02	fast GS		
×	Détails	28 janv. 2025, 16:20:38	fast GS		

Nom

Modifié le

Typ

Do:

vib-bits

+ Redémarrer l'application !

Réconcilier ses données

Qu'est ce que la réconciliation ?

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

Victor Hugo (Q535)

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

France (Q142)

pays de nationalité (P27)

<https://www.wikidata.org/wiki/Property:P27>

↑
↓
Référentiel auteur de la BNF
VIAF
etc

↑
↓
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

			cd_nom	nom_scientifique	nom_vernacula	famille
☆	🗨	1.	67778	Facette	truite de rivière	Salmonidae
☆	🗨	2.	67854	Filter 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	crustivomer	Salmonidae
☆	🗨	6.	67759	Trier...	ombre commun	Salmonidae
☆	🗨	7.	67794	Aperçu	luchon	Salmonidae
☆	🗨	8.	67765	Réconcilier	saumon atlantique	Salmonidae
☆	🗨	9.	67111	Démarrer la réconciliation...	ablette	Cyprinidae
☆	🗨	10.	67466	Extract named entities...	Facettes	
				Actions		
				Copier les données de réconciliation...		

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 la colonne "non

Services

Wikidata Reconciliation for
OpenRefine (en)

Ajouter un service standard...

Indiquer l'URL du service:

<https://tools.wmflabs.org/openrefine-wikidata/en/api>

Ajouter un service

Annuler

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

☐ Réconcilier sans type particulier

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

Nombre maximal de candidats renvoyés

Réconcilier ses données


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

Equivalent d'une coche simple

Apparier cette cellule

Apparier toutes les cellules identiques

Annuler



Victor Hugo (Q535)
French poet, novelist, and dramatist

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

Réconcilier ses données

Pour récupérer des données à partir de wikiData

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

Ex : date of birth

Ex: child

Add Property

Preview

date of birth|

Select an item from the list:

date of birth P569

nom	date of birth
Romain Gary Choisir une nouvelle correspondance	1914-05-21T00:00:00Z
Victor Hugo Choisir une nouvelle correspondance	1802-02-26T01:00:00Z

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