Récupération et intégration de lots de données depuis le GBIF

Fonctionnement global

1. Configurer et déclencher un export ciblé via l'API du GBIF
2. Récupérer le fichier généré sur le serveur
3. Préparer les données et les stocker dans la base de données
4. Préparer les métadonnées associées
5. Intégrer ou actualiser les données dans la synthèse

Scripts

R : Préparer l'export GBIF

La récupération d'un export de données pré-filtré selon différentes variables sélectionnées, et pour des grands volumes de données, est possible en configurant un export via l'API download du GBIF (par R par exemple comme décrit ici, ou en python). Cette "commande" consiste à transmettre une requête au GBIF, qui met à disposition un fichier de données correspondant aux filtres configurés, dans l'espace du compte qui a commandé la requête. Il est donc nécessaire d'avoir un compte GBIF actif.

Dans le fonctionnement du pôle invertébrés, les données filtrées concernent : * Les données disposant de coordonnées géographiques * Les données sans problème géospatial (géométries invalides etc) * Les données disponibles en Auvergne-Rhône-Alpes * Les données d'un ou certains jeux de données ciblés à l'avance * Les données de faune invertébrée des groupes les mieux connus

Pour connaitre les "PHYLUM_KEY" à filtrer, les correspondances avec taxref sont disponibles par défaut dans les bases de données GeoNature dans taxonomie.taxref_liens. Par exemple pour les invertébrés :

```
SELECT DISTINCT t.cd_nom, tl.ct_sp_id
FROM taxonomie.taxref t
JOIN taxonomie.taxref_liens tl ON tl.cd_nom = t.cd_nom
WHERE tl.ct_name ='GBIF'
AND t.id_rang = 'PH'
AND t.regne='Animalia'
AND t.phylum!='Chordata';
```

À titre d'exemple pour récupérer les Annélides, Mollusques et Arthropodes du jeu de données iNaturalist, la configuration de l'export est la suivante :

```
{
 "creator": "USER",
```

```
 "notificationAddresses": [
 "EMAIL"
],
"sendNotification": true,
"format": "DWCA",
"predicate": {
"type": "and",
"predicates": [
 {
  "type": "equals",
  "key": "HAS_COORDINATE",
  "value": "True"
 },
 {
  "type": "equals",
  "key": "HAS_GEOSPATIAL_ISSUE",
  "value": "False"
 },
 {
  "type": "equals",
  "key": "GADM_GID",
  "value": "FRA.1_1"
 },
 {
  "type": "in",
  "key": "DATASET_KEY",
  "values": [ "50c9509d-22c7-4a22-a47d-8c48425ef4a7" ]
 },
 {
  "type": "in",
  "key": "PHYLUM_KEY",
  "values": [ "42", "52", "54" ]
 }
]
}
}
```

Serveur : Récupérer, préparer et stocker les données en base

Postgresql : Insérer ou actualiser les données en synthèse

From:
https://wiki-sinp.cbn-alpin.fr/ - **CBNA SINP**

Permanent link:
https://wiki-sinp.cbn-alpin.fr/procedures/recuperation_et_integration_de_donnees_depuis_le_gbif?rev=1769897308

Last update: **2026/01/31 22:08**