Import des données de test

Récupération du dépôt "data"

- En local, sur votre machine, récupérer le dépôt Github "sinp-paca-data" : `git clone git@github.com:cbn-alpin/sinp-paca-data.git`
- Suivre les éventuelles étapes indiquées dans les fichiers README.md (CBNA) et README.md (CEN-PACA)
- Sur le serveur, sauvegarder le dossier "data" : voir ci-dessous
- Se placer à la racine du dossier *sinp-paca-data*
- Transférer les scripts : `rsync -av --copy-unsafe-links --exclude var --exclude .git --exclude .editorconfig --exclude .vscode --exclude .gitignore --exclude settings.ini --exclude "data/raw/*" --exclude venv ./ geonat@db-paca-sinp:~/data/ --dry-run`
    - Supprimer l'option `--dry-run` si tout semble ok pour effectuer le transfert réel
- Se connecter au serveur
    - Créer les fichiers *settings.ini* à partir des fichiers *settings.sample.ini* pour : *cbna*, *cenpaca* et *shared*
    - Préparer l'environnement du script *import-parser* en suivant les indications du fichier README.md (import-parser)
        - **Notes** : il est nécessaire de redonner les droits d'execution à GCC pour tout le monde si l'on veut pouvoir installer correctement le venv avec `sudo chmod o+x /usr/bin/gcc`. Une fois l'installation terminée, retirer les à nouveau avec '' chmod o-x /usr/bin/gcc ''.
- **Notes** : les données brutes nécessaires aux scripts sont automatiquement téléchargées depuis Dropbox

Sauvegarde du dossier "data"

- Se connecter à l'utilisateur "admin" sur le serveur : `ssh geonat@db-paca-sinp`
- Créer le dossier de sauvegarde : '' mkdir -p backups/data ''
- Sauvegarder le dossier "data" : '' mv data backups/data/$(date +%F) ''
- Recréer le dossier "data" vide : '' mkdir data ''

Exécution du script d'import dans GeoNature module Synthèse

- Se connecter au serveur "db" en tant qu'admin : `ssh geonat@db-paca-sinp`
- Se placer dans le dossier // ~/data/cbna/ // de l'utilisateur *geonat* : `cd ~/data/cbna`
- Créer le fichier de configuration : `cp config/settings.sample.ini config/settings.ini`
- Modifier le fichier *settings.ini* en fonction de l'installation locale : `vi config/settings.ini`
- Nettoyer le dossier *data/raw/* si nécessaire : '' cd ~/data/cbna/data/raw/ ; rm -f *.{csv,ini} ''
- Lancer une session détachable : `screen -S import-cbna`

    - Voir la documentation générale concernant les commandes à utiliser avec Screen pour quitter puis se reconnecter à une session.
- Lancer le script : ''cd bin/ ; ./import_initial.sh -v ''
- Vérifier l'absence d'erreurs dans la session et vérifier le contenu de la base de données

Mise à jour des données GeoNature Atlas

- **Ressources** : https://github.com/PnX-SI/GeoNature-atlas/blob/master/docs/vues_materialisees_maj.rst
- Au préalable, ajouter une authentification HTTP au site nature.silene.eu, le temps des tests avant mise en prod.
- Se connecter au serveur : `ssh admin@<sinp-db>`
- Lancer une session détachable : `screen -S import-cbna`
    - Voir la documentation générale concernant les commandes à utiliser avec Screen pour quitter puis se reconnecter à une session.
- Lancer la commande : `psql`
    - Sélectionner la base de l'atlas : `\c gnatlas`
    - Exécuter les requête suivante suivant le besoin
        - Pour mettre à jour uniquement les données d'observation (table *synthese*) : SELECT `atlas.refresh_materialized_view_data() ;`
        - Si le territoire du SINP à changé et qu'il faut remettre à jour le référentiel géographique : '' SELECT atlas.refresh_materialized_view_ref_geo() ''

Exécution du script d'import des images de l'INPN

- Se connecter sur le serveur : `ssh geonat@sinp-<region>-web`
- Se placer dans le dossier du script dans TaxHub : `cd ~/www/taxhub/data/scripts/import_inpn_media '' * Créer un environnement virtuel : '' python3 -m venv venv '' * Activer l'environnement virtuel : '' source venv/bin/activate '' * Redonner les droits d'exécution à GCC pour tout le monde si l'on veut pouvoir installer correctement les paquets Python dans le venv : '' sudo chmod o+x /usr/bin/gcc '' * Installer les paquets suivant : '' pip install psycopg2 requests '' * Retirer les droits d'exécution à GCC pour tout le monde : '' sudo chmod o-x /usr/bin/gcc ''. * Créer le fichier de configuration : '' cp config.py.sample config.py '' * Modifier les paramètres : * '' SQLALCHEMY_DATABASE_URI = "postgresql://geonatadmin:<mot-de-passe>@10.0.1.10:5432/geonature2db"`
    - `QUERY_SELECT_CDREF = """SELECT DISTINCT cd_ref FROM taxonomie.bib_noms ORDER BY cd_ref LIMIT 100"""` : supprimer le '' LIMIT 100 '' une fois un premier test effectué
- Lancer le script : '' python import_inpn_media.py ''
- Si tout c'est bien passé, désactiver l'environnement virtuel : '' deactivate ''
- Le script ajoute les photos en tant que "secondaire" pour désigner celle avec le plus petit id_media (= au hasard) comme "principale", utiliser la requête :

```
WITH first_media AS (
    SELECT MIN(id_media) AS first_id_media_founded, cd_ref
```

```
    FROM taxonomie.t_medias
    GROUP BY cd_ref
)
UPDATE taxonomie.t_medias AS tm
    SET id_type = 1
    FROM first_media AS fm
    WHERE tm.id_media = fm.first_id_media_founded
        AND tm.cd_ref = fm.cd_ref ;
```

- Pour afficher les images sur l'Atlas, il est nécessaire de rafraichir les données des vues matérialisées *atlas.vm_medias* et *atlas.vm_taxons_plus_observes* :

```
REFRESH MATERIALIZED VIEW atlas.vm_medias WITH DATA ;
REFRESH MATERIALIZED VIEW atlas.vm_taxons_plus_observes WITH DATA ;
```

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

Permanent link:
https://wiki-sinp.cbn-alpin.fr/database/sinp-paca/import-test

Last update: **2022/04/20 09:13**