Réflexions/Discussions sur le format d'échange

- Voir si on utilise plutôt un chaine vide sans guillemet pour indiquer une valeur NULL dans le format d'import
- A priori, dans le format d'import, l'utilisation d'une tabulation (qui ne se rencontre pratiquement jamais dans les valeurs des champs) doit permettre d'éviter l'utilisation des guillemets pour encapsuler une valeur de champ même si celui-ci en contient.
- Lors de l'import avec *COPY* il n'y a pas de vérification des doublons. Si nous répétons la procédure d'import les méta-données peuvent être importé plusieurs fois. Solutions :
    - le script d'import doit se charger de réinitialiser ces tables avant chaque import s'ils trouvent les codes/UUID qui vont être importés via les fichiers CSV
    - pour les métadonnées nous pouvons peut être utiliser un script SQL plus classique qui se chargerait d'éviter l'import si les données sont déjà présentes. Inconvénients : moins lisible, 2 formats et traitements différents. Avantages : gestion complètes des méta-données (protocoles, acteurs de type personne...).
    - le script d'import via *COPY* charge les données dans une table temporaire (copie vide de la table de destination) puis les injectes dans la table de destination en s'assurant qu'il n'y a pas de doublon (*NOT EXISTS*).
- Prévoir un fichier pouvant indiquer la version du format des fichiers ci-dessous. Cela implique de versionner ce format.
- Ne devrait on pas transférer les champ *code_nomenclature_blurring* et *code_nomenclature_diffusion_level* dans DATASET ? Avantage : cela aurait plus de sens vis à vis de notre décision de gérer par jeu de données complet ces infos. Inconvénient : oblige à transférer un fichier *dataset.csv*, rend plus difficile l'intégration des infos dans GeoNature...

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

Permanent link:
https://sinp-wiki.cbn-alpin.fr/discussion/database/import-formats?rev=1598860218

Last update: **2020/08/31 07:50**