

Versement de données dans la base de données CETTIA (V1)

Notice de préparation des données en vue d'un import à l'attention des producteurs de données

Préambule : Une alternative est la saisie des données directement dans l'outil de saisie de CETTIA. Dans ce cas, rendez-vous sur la page d'accueil du portail de saisie (<http://cettia-idf.fr/bdd>) pour créer un compte. Vous pouvez vous rapprocher de l'agence régionale de la biodiversité (ARB, anciennement Natureparif) pour toute information relative à la saisie (des formations à la prise en main de l'outil sont également régulièrement organisées). **Contact** : sinp@iau-idf.fr

Format et préparation des données

Afin de permettre un import efficace de vos données dans la base de données régionales CETTIA, il est nécessaire de les préparer sous la forme d'un fichier au format Excel ou autre tableur compatible. Les colonnes du fichier excel à importer doivent être renommées selon leur contenu. Certaines informations doivent être obligatoirement fournies, d'autres sont facultatives. Le fichier ne doit pas contenir plusieurs colonnes avec le même nom. Les colonnes ne correspondant pas à un libellé prévu pour les fichiers d'imports sont ignorés.

A - Champs obligatoires

1 - identification de l'origine des données

a) **nom_base_origine**

> texte libre - max 48 caractères.

b) **id_origine**

Identifiant unique de la donnée (non obligatoire mais fortement recommandé).

> chiffre et lettre non accentués, tirets - max 64 caractères

2 - identifiant de l'observateur

a) **mail_observateur**

Meilleure solution pour identifier un compte sans ambiguïté. Il faut que ce soit l'adresse utilisée pour le compte de la personne concernée.

OU

b) **nom_observateur** ET **prenom_observateur**

OU

c) **nom_observateur_complet** avec nom + prénom ou prénom + nom, etc...

Les différents identifiants (nom et prénom etc...) pour les observateurs seront reliés à un compte via l'interface. Si le compte n'existe pas, il pourra être créé.

3 - Identifiant du taxon

a) **cd_nom_taxref**

Identifiant TAXREF

OU

b) **nom_scientifique**

Nom scientifique complet (avec auteur)

OU

c) **libel_nom**

Nom scientifique sans auteur

> Si le **cd_nom_taxref** est renseigné, il peut-être utile de préciser également un **nom_scientifique** ou un **libel_nom**, afin de conserver une trace du taxon d'origine en toutes lettres.

4 - Date

a) **date_complete**

> Différents formats possibles: 1/1/2015 ou 01-01-2015 ou 1 janv. 2015 ou 1er janvier 2015 ou 1-1-15, etc. (pour année sur 2 chiffres, si inférieur à 50 interprété comme 2000, sinon 1900).

OU

b) **jour ET mois ET annee**

> Uniquement nombres

OU

b) **date_format_sql**

> Dates au format sql (« 2016-11-15 » par exemple – l'heure peut être incluse : « 2016-11-15 09:28:00 » par exemple).

Les dates imprécises ne peuvent pas être mentionnées avec ce format.

Pour indiquer une date imprécise, plusieurs possibilités :

a) Indiquer une **date_complete** sans le jour (08-2015 par exemple) qui sera interprété comme « en août 2015 », ou bien avec seulement l'année (les données seront alors datées par une période allant du 1^{er} janvier au 31 décembre).

b) ajouter une colonne **date_max** qui indiquera le dernier jour de la période alors que **date_complete** indique le premier jour de la période.

> Cette possibilité est à réserver à des cas bien particuliers, notamment pour les données historiques.

5 - Quantité

a) **nombre_total**

Doit être un nombre. Une colonne vide ou contenant autre chose qu'un nombre sera interprété comme "présence". 0 est interprété comme absence remarqué.

6 - Commune

a) **insee**

Le n° insee de la commune. C'est la solution la plus sûre.

OU

b) **code_postal ET nom_commune**

OU

c) **num_dept ET nom_commune**

Note: si le fichier contient une colonne avec le nom de la commune et un code postal, un numéro insee, ou

un numéro de département, il suffit de dupliquer la colonne et de la renommer **code_postal** ou **insee**. Il est inutile de supprimer le nom de la commune, les caractères autre que les chiffres seront automatiquement ignorés.

7 - Coordonnées géographiques

a) **latitude** ET **longitude**

Coordonnées géographiques WGS84 en décimal ou sexagésimal
OU

b) **lambert_93_x** ET **lambert_93_y**

Coordonnées en lambert 93

8 - précision des coordonnées

La colonne **precision_localisation** doit contenir une des valeurs suivantes:

- **10, 20, 50 ou 100** pour une localisation précise (en mètres).
- **200** pour une localisation imprécise type cettia
- **commune** pour une localisation à la commune, pour des données historiques
- **2km** ou **10km** pour des données historiques.

B - Champs facultatifs

9 – La structure

nom_structure

Nom permettant d'identifier une structure. Les différents identifiants (noms de structures) seront chacun reliés depuis l'interface à une structure enregistrée dans la base. Si la structure est inconnue dans la base de donnée, elle pourra être créée.

10 – L'étude ou le projet

nom_etude

Permet de lier des données à une étude ou un projet précis. Les caractéristiques de l'étude doivent impérativement avoir été configurées préalablement. S'il s'agit d'une étude réalisée par une autre structure que celle mentionnée précédemment (par exemple un bureau d'étude), cette autre structure pourra être mentionnée lors de la création de l'étude depuis l'interface Cettia.

11 – Détail sur l'effectif

precision_effectif

Précision pour l'effectif total. Valeurs possibles : ">", "~" ou "≈". Ces 3 valeurs peuvent aussi être fournies dans la colonne **nombre_total**, devant la valeur numérique.

nb_males

nb_femelles

nb_males_ad

nb_males_immatures

nb_males_juv

nb_femelles_ad

nb_femelles_immatures

nb_femelles_juv

nb_ad

nb_imm

nb_juv

nb_type_femelle

nb_larves

nb_exuvies

> Ces colonnes doivent contenir uniquement des nombres. Si autre chose, le contenu non vide sera uniquement utilisé pour ajouter les éventuels mots-clés correspondant (« adultes », « juvéniles », « larves », ...).

abondance_bota

Valeurs possibles : R pour « rare, qqs pieds », I pour « Un individu », + pour « <1%, peu abondant' », S1 pour « < 5% », S2 pour « 5-25% », S3 pour « 25-50% », S4 pour « 50-75% », S5 pour « 75-100% », A pour « 1 à 10 pieds », B pour « 11 à 100 », C pour « 101 à 1000 », D pour « 1001 à 10000 », E pour « > 10000 »

12 – Heure de l'observation

Colonne **heure**

> Format 10:04 ou 10h04

13 – remarques et commentaires

> Toutes les colonnes ayant un nom commençant par **remarque** (par exemple **remarque1**, **remarque2**, **remarque3**, etc.) seront concaténées en un seul champ commentaire. Au total, 1000 caractères max (sinon sera tronqué).

14 – Mots-clés

La colonne **mots_clefs** contient un ensemble de texte pouvant correspondre à des mots-clés utilisés dans Cettia. Chaque texte (un ou quelques mots) doit être séparé des autres par le caractère #. Cette colonne peut être construite avec la fonction concaténée dans Libre Office ou Excel. Ce n'est pas un problème si deux mots-clés sont séparés par plusieurs caractères #.

15 – Données complémentaires

Il est possible dans certains cas d'importer des informations annexes, en particulier pour les protocoles. Cette solution nécessite cependant des interventions manuelles dans la base de données. Toutes les colonnes préfixées **data_** seront enregistrées dans une table annexe. Le nom de la colonne doit correspondre au contenu de la colonne. Par exemple, on nommera la colonne **data_plage_distance** pour les plages de distance d'observation depuis un point d'écoute oiseaux.