

## Extractions Données

L'utilisation du package `extractionsDonnees` nécessite l'installation de R et de RStudio :

R : <https://cran.rstudio.com/>

RStudio : <https://posit.co/download/rstudio-desktop/>

Il est ensuite possible d'utiliser le script R déjà présent dans le dossier

"~BDD/EXPORT\_DONNEES/SCRIPT\_EXPORT" ou de créer une copie dans son propre dossier de travail.

Si le NAS n'est pas accessible, il est possible de télécharger le package à partir de ce lien :

<https://github.com/berenicegc/extractionsDonnees>, en cliquant sur le fichier

« `extractionsGeoNature_0.0.0.9000.tar.gz` » puis en cliquant sur l'icône de téléchargement.

Après le téléchargement, un script R doit être créé, sans oublier d'installer et de charger le package :

```
install.packages("/Volumes/BDD/EXPORT_DONNEES/SCRIPT_EXPORT/extractionsDonnees_0.0.0.9000.tar.gz", repos = NULL, type = "source")
```

```
library(extractionsDonnees)
```

## Import des fichiers nécessaires

La fonction **`import`** permet d'importer les fichiers nécessaires à l'extraction des données contenues dans les champs additionnels du fichier. Deux fichiers sont nécessaires : les données téléchargées sur la plateforme GéoNature et le fichier TaxRef.

### Marche à suivre

1. Télécharger les données en .csv sur GeoNature
2. Mettre le csv dans le même dossier que le script R utilisé ou spécifier le dossier dans lequel les fichiers sont présents dans l'argument 'path', le chemin devant être sous la forme "~Downloads". Dans tous les cas, le fichier `synthese_observations` de GeoNature et TaxRef doivent être dans le même dossier.

Par défaut, si l'argument 'path' n'est pas précisé, les fichiers seront cherchés dans le même dossier que le script R.

Si plusieurs synthèses GeoNature sont présentes dans le même dossier, c'est le fichier le plus récent (téléchargé en dernier) qui sera importé, pour éviter l'ouverture de tous les fichiers en simultané et donc potentiellement la présence de doublons.

3. Préciser le nom des fichiers s'ils ont été changés avec les arguments 'file' et 'TaxRef'. Si les noms des fichiers n'ont pas été modifiés il n'est pas nécessaire de préciser ces arguments et les noms par défaut des fichiers téléchargés seront pris en compte. Attention, les noms de fichiers ne doivent pas contenir d'accents ou de caractères spéciaux. Le nom des fichiers peut être écrit avec ou sans l'extension (.csv), toutefois les fichiers doivent obligatoirement être des .csv (et non des .xls ou .xlsx). Si l'import ne marche pas, relancer la fonction peut fonctionner.

## Exemples

Import par défaut (fichiers non renommés et présents dans le même dossier que le script R)

```
import()
```

Import avec modification du chemin d'accès

```
import(path = "~/Documents")  
import(path = "~/Downloads")
```

Import avec modification des noms de fichiers (les fichiers doivent être des .csv)

```
import(file = "BDD_ABAURA.csv", TaxRef = "TaxRefv16.csv")
```

## Extraction des colonnes d'intérêt

La fonction **extract** permet d'extraire les colonnes d'intérêt, contenues dans les champs additionnels de la synthèse GéoNature. Les colonnes peuvent être : plante, caste, station, année de détermination et méthode de capture. Un tri des données de plantes pourra également être réalisé dans la fonction.

### Marche à suivre

1. Sélectionner les colonnes à extraire (plante, caste, station, annee\_determination, methode) à l'aide de l'argument 'col.' Il est possible de ne sélectionner qu'une colonne ou plusieurs, l'ordre n'ayant pas d'importance.

Si une seule colonne est sélectionnée, la fonction s'écrit sous la forme `extract(col = "plante")`.

Dans le cas où plusieurs colonnes sont sélectionnées, la fonction prend la forme `extract(col = c("plante", "caste", "annee_determination"))` par exemple, `c()` permettant de combiner des valeurs.

Par défaut, si l'argument "col" n'est pas précisé, toutes les colonnes sont sélectionnées.

2. Sélectionner le degré de précision taxonomique choisie ("famille", "genre" ou "sp") à l'aide de l'argument "precision\_taxo". Si un rang taxonomique plus grossier est sélectionné (famille ou genre), les rangs plus précis seront conservés. Par exemple, si

on autorise de conserver une précision allant jusqu'à la famille, les observations allant jusqu'au genre ou à l'espèce seront conservées.

Par défaut, toutes les observations sont conservées et aucun tri n'est effectué.

La fonction `extract` peut simplement s'écrire `extract()`, dans ce cas toutes les colonnes utiles des champs additionnels seront extraites et toutes les données de plantes seront gardées (même les observations NA ou les données ne renvoyant pas à un `cd_nom` dans `TaxRef`).

Il faut compter environ 2 minutes d'attente pour l'export de toutes les colonnes et pour une base de données d'environ 100 000 observations

### Exemples

Extraction par défaut (toutes les colonnes des champs additionnels et sans mise en place d'un filtre sur la précision taxonomique)

```
extract()
```

Extraction avec sélection de colonnes

```
extract(col = "plante")  
extract(col = c("plante", "caste", "station",  
"annee_determination", "methode"))
```

Extraction avec filtre par précision taxonomique (la colonne 'plante' doit obligatoirement être sélectionnée pour préciser cet argument)

```
extract(col = "plante", precision_taxo = "sp")
```

### Export du fichier final

La fonction 'export' permet d'exporter la base de données finale, contenant les variables sélectionnées.

Par défaut, le fichier est appelé "export\_final\_donnees.csv" et est stocké au même endroit que le script R.

Il est possible de changer le nom du fichier à l'aide de l'argument 'fichier', en spécifiant l'extension .csv (par exemple : fichier.csv).

Il est également possible de changer le chemin d'accès à l'aide de l'argument 'path', qui doit être sous la forme «~/Documents» ou «~/Downloads», comme pour la fonction 'import'.

### Exemples

Export par défaut (fichiers non renommés et présents dans le même dossier que le script R)

```
export()
```

Export avec modification du chemin d'accès

```
export(path = "~/Documents")
```

```
export(path = "~/Downloads")
```

Export avec modification des noms de fichiers (les fichiers doivent être des .csv)

```
export(fichier = "BDD_ABAURA.csv")
```