## Introduction

Pour faciliter le travail NSU, nous mettons à votre disposition plusieurs programmes qui vous serviront à :

* Télécharger l’ensemble des bases du serveur de manière automatisée
* Fusionner les multiples bases NSU en un nombre réduit de bases
* Identifier des données invraisemblables ou incohérentes dans les bases
* Classer les images

## Consignes générales

Pour garantir le bon fonctionnement des programmes et, le cas échéant, un bon appui de dépannage, il est fortement déconseillé de modifier la moindre ligne de code. Sont exempts de cette règle générale les fichiers qui contiennent les paramètres (e.g., chemin du répertorie principal, nom du pays, mot de passe, etc.), lesquels doivent subir un ajustement pour faire fonctionner les programmes.

En dépit du bon respect de cette consigne, il est possible que les programmes ne fonctionnent pas comme prévu. Pour anticiper certains problèmes, ce guide prévoie des astuces de dépannage. Pour les problèmes qui ne sont pas traités dans ce guide, veuillez-vous adresser à un membre de l'équipe CAPI qui vous appuie.

Lors de cet échange, pour nous aider à vous aider, veuillez fournir les informations suivantes :

* **Description du problème**. Pour être utile, une description doit aller au-delà de "votre programme ne marche pas". Au lieu de cela, essayer de nous donner autant d'information que possible. Par exemple : des captures d’écran, des fichiers logs, etc.
* **Etapes pour reproduire le problème.** Afin de comprendre puis corriger le problème, nous devons être en mesure de le reproduire. Pour ce faire, nous avons besoin des étapes à suivre pour reproduire le problème que vous avez identifié. Parfois, nous pouvons avoir besoin, en plus de cela, d'un programme ou d'un extrait de données pour nous mettre dans le contexte de votre problème.

## Inventaire des fichiers, et explication brève de leur fonctionnement

Le fichier « programmes NSU.zip » un système de répertoires, et certains fichiers à l’intérieur de ces répertoires. Voici un inventaire ce qui vous est transmis :

* **donneesBrutes.** Répertoire où loger les données brutes. Sont destinés à ce répertoire deux types de fichiers. Pour le premier, il s’agit des fichiers zip téléchargés depuis votre serveur, et qui contiennent des données en format Stata ou Tabular. Pour le second type il s’agit des données qui sont décomprimées à partir des fichiers zip précédemment cités par le programme de fusion décrit plus bas. Sont exclus de répertoire tout autre fichier « étranger »--par exemple, des fichiers SPSS (ou fichier zip qui contient des bases SPSS). Notez bien : ne décomprimez pas les fichiers zip ; le programme de fusion accomplit cette tâche et classe les fichiers décomprimés.
* **donneesFinales.** Répertoire où le programme de fusion sauvegarde les fusionnées à partir des multiples bases de l’enquête NSU.
* **donneesProvisoires.** Répertoire où les bases de données temporaires seront logées au cours de l’exécution du programme de fusion.
* **Etiquettes.** Répertoire où les étiquettes de valeur des produits et des unités sont rassemblées et combinées.
* **imagesBrutes.** Répertoire où loger les fichiers zip téléchargés depuis votre serveur qui contiennent des images. Notez bien : ne décomprimez pas les fichiers zip ; le programme de classement des images accomplit cette tâche et classe les fichiers décomprimés.
* **imagesClassees.** Répertoire où les images sont reclassées d’abord par produit et ensuite par unité de ce produit.
* **logs.** Répertoire où se trouve le log des programmes Stata décrits plus bas, ainsi que des fichiers qui captent et décrivent d’éventuelles erreurs produites lors de l’exécution des programmes.
* **ressources.** Répertoire qui contient des ressources utilisées pendant l’exécution des programmes. Les fichiers Excel compilent les produits, les unités, les tailles admises, ainsi que leurs codes. Le fichier .doi contient une syntaxe pour raccourcir ou autrement changer le nom de certains répertoires créés avec le programme de classement des images. Le fichier .do crée des bases Stata à partir des bases tabular, au cas où l’utilisateur dispose d’une version de Stata antérieure à la version 14.
* **downloadAll\_vf.R.** Programme R pour télécharger toutes les bases d’un certain format depuis le serveur désigné. Le programme est décrit en plus de détails plus bas.
* **serverDetails.R.** Programme R qui rassemble et capte les détails du serveur de l’utilisateur.
* **lancer\_FusionnerBasesNSU.do.** Programme Stata qui capte certains paramètres et lance la fusion des bases.
* **fusionnerBasesNSU\_v1\_INS.do.** Programme lancé par le programme précédent.
* **lancer\_classerNsuImages.do.** Programme Stata qui capte certains paramètres de l’utilisateur et lance le reclassement des images dans un nouveau système de répertoires.
* **classerNsuImages\_v1\_INS.do.** Programme lancé par le programme précédent.

## Télécharger toutes les bases (downloadAll\_vf.R)

### Fonctionnement

Télécharge toutes les bases du serveur pour un format de données spécifié.

### Description des composants

Constituté de trois composants :

1. **downloadAll\_vf.R .** Compose le cœur du programme--c'est à dire, la logique de requête de donnée auprès du serveur et d'analyse de réponse du serveur.
2. **serverDetails.R**. Regroupe les paramètres que l'utilisateur doit spécifier : l'adresse du serveur, l'authentification (les références) pour y accéder, le format de données désiré, et le répertoire où stocker les bases téléchargées.
3. **Programmes de service**. Comme la majeur partie de programmes R, ce programme emploie des programmes de service, des packages dans le jargon de R, qui sont téléchargés et installés, au besoin, au cours de l'exécution du programme downloadAll\_vf.R . Sans ces programmes l'exécution du programme va se solder en échec.

### Comment installer

#### Installer R

R est un programme statistique gratuit. Pour l'installer, suivre ces étapes

Naviguer vers le site du projet : <https://cran.r-project.org/>



Cliquer sur Download R for Windows



Cliquer sur base



Sélectionner Download R 3.X.X for Windows

Installer le logiciel

#### Installer R Studio

R Studio est une interface pour R gratuite qui facilite l'exploitation de R. Pour l'installer, suivre ces étapes :

Naviguer vers le site : <https://www.rstudio.com/products/rstudio/download/>



Cliquer sur le bouton DOWNLOAD la version gratuite de R Studio



Cliquer sur la version pour votre système d'exploitation, très probablement Windows Vista/7/8/10

Installer le logiciel

#### Paramétrer le programme

Avant de lancer le programme, un certain paramétrage est requis. Voici les étapes à suivre :

##### Créer un utilisateur API sur votre serveur



Un utilisateur API est un utilisateur spécial qui peut accomplir ses actions à travers l'Application Programming Interface (API) de Survey Solutions. Aucun autre utilisateur, sauf l'admin, ne peut agir ainsi.

Pour ce faire, se connecter au serveur en tant qu'utilisateur admin. Du menu Équipes et rôles, sélectionner Utilisateur de l'API. Créer un utilisateur en suivant les étapes habituelles, et en veillant à garder le nom d'utilisateur et mot de passe pour la suite

##### Ouvrir R Studio



##### Ouvrir serverDetails.R

Avec R Studio, ouvrir le fichier serverDetails.R



##### Modifier chaque paramètre de serverDetails.R

La valeur des paramètres doivent prendre la place de XXX (valeur de défaut) et rester à l'intérieur des guillemets.



Voici une explication plus détaillée que dans les commentaires du programme serverDetails.R des paramètres à spécifier :

* **server**. Partie de l'adresse qui suit "https://" et précède ".mysurvey.solutions". Si l'adresse de mon serveur est "https:// *uemoa*.mysurvey.solutions" , la valeur du paramètre server devrait être "uemoa"
* **login**. Nom d'un utilisateur API.
* **password**. Mot de cet utilisateur.
* **dataType**. Format de données demande du serveur. Ce format vaut pour l'ensemble des bases téléchargées suite à l'exécution du programme. Ce paramètre peut être changé avant de lancer le programme à nouveau afin d'obtenir toutes les bases dans un autre format. Noter que les seules valeurs admissibles sont : Tabular, STATA, et SPSS. Le format Binary est également admissible pour obtenir les photos, mais son utilisation n'a pas été suffisamment testée et demande une très bonne connexion pour obtenir ces fichiers volumineux.
* **dataDir**. Adresse sur votre machine où télécharger les bases lors de l'exécution de ce programme. Deux consignes à suivre. Premièrement, utiliser le slash, / , au lieu de l'antislash, \ , dans les chemins de répertoire, même sur les machines Windows. R s'attend à des chemins uniquement avec des slashs. Deuxièmement, veiller à mettre un slash à la fin du chemin du répertoire. Si vous copier et coller le chemin sur une machine Windows, le slash terminal manquera et devra être composé manuellement

### Comment utiliser

Pour lancer le programme,

##### Vérifiez l'existence d'une connexion internet

Ceci est nécessaire pour communiquer avec le serveur et télécharger les bases.

##### Ouvrir R Studio

Voir image dans la section précédente

##### Ouvrir downloadAll\_vf.R

Avec R Studio, sélectionner File > Open File dans le menu du programme et en sélectionnant le programme downloadAll\_vf.R

##### Lancer le programme

Cliquer sur Source en haut et droit du panneau de l'éditeur de R Studio



Vous saurez que le programme s'est lancé par deux signes. Premièrement, vous verrez une ligne de commande décrivant votre programme comme "source." Deuxièmement, vous verrez, après une petite période d'attente, des textes paraître décrivant le processus du programme--d'abord l'installation de "packages", ensuite le lancement de requêtes de bases pour chaque masque importé au serveur.



Pour télécharger les bases dans un autre format, modifier le paramètre dataType, et relancer le programme à nouveau. Cette exécution téléchargera l'ensemble des bases dans ce format nouvellement spécifié.

### Comment dépanner

Afin de dépanner, il faut lire le message d'erreur qui s'affiche.

Voici les solutions aux problèmes connus.

* **Mauvais paramétrage.** Le programme s'attend à avoir certain paramètres renseignés et des valeurs valides pour ces paramètres. Les messages d'erreur 1a à 1f concernent les problèmes de paramétrage, et décrivent bien la nature du problème
* **Echec d'installation d'un "package".** Lors du premier lancement du programme, R installe certains "packages", des programmes de service, qui sont nécessaires pour le bon déroulement du programme. Lors de ce processus, il est possible que certains de ces packages ne s'installent pas correctement. Si une erreur concernant un "package" s'afficher, essayer d'installer les packages manuellement. S'il est possible d'identifier le "package", installez seulement celui-là. S'il n'est pas possible de l'identifier, installer l'ensemble des "packages" requis. Pour ce programme, voici les packages requis : httr, RCurl, stringr, et lubridate. Pour installer un seul "package", par exemple httr, utiliser la syntaxe suivante : install.packages("httr") , où l'on remplace httr par le "package" manquant. Pour installer tous les "packages", utiliser la syntaxe suivante : install.packages(c("httr", "RCurl", "stringr", "lubridate"))
* **Impossible de télécharger les bases de certains masques avec le programme.** Pour divers raisons, il est possible que certaines bases ne puissent pas être téléchargées. Pour documenter les problèmes, le programme capte ces problèmes dans le fichier "échecs de téléchargement.csv". Les observations de ce fichier correspondent aux bases non-téléchargées, avec mention du titre du masque, version, type de données demandé, motif d'échec, et l’emplacement dans le programme de l'échec. Pour obtenir ces bases, il est conseillé de les télécharger manuellement en connectant au serveur avec un utilisateur Headquarters.
* **Problème de connexion au serveur dû aux problèmes de réseau/WiFi.** Après un temps d'attente sans connexion, le programme échouera. Si la connexion internet revient, essayer de lancer le programme à nouveau. Si de tels problèmes persistent, veuillez-vous adresser à l'équipe d'appui de la Banque Mondiale.

## Fusionner les bases (lancer\_FusionnerBasesNSU.do)

### Fonctionnement

Fusionne les nombreuses bases de l'enquête NSU.

Une description plus longue des tâches accomplies par ce programme :

* Vérifie le paramétrage
* Décomprime les archives de données brutes
* Harmonise le nom de variables dans l’ensemble des bases
* Crée des bases provisoires avec des noms de variables harmonisés
* Extrait les étiquettes de valeur des produits et d’unité afin de
* Joindre toutes les bases du même type
* Intègre dans ces bases les identifiants de marché
* Intègre ou crée, au besoin, un identifiant de relevé
* Applique les étiquettes de valeur aux identifiants de
* Sauvegarde trois bases dans le répertoire donneesFinales

### Description des composants

Composé des fichiers suivants :

* **lancer\_FusionnerBasesNSU.do**. Capte les paramètres à spécifier par l'utilisateur. Également, comme son nom l'indique, il lance le cœur du programme.
* **fusionnerBasesNSU\_v1\_INS.do**. Contient les instructions pour fusionner et mettre en forme les bases
* **produitUniteRaccourci\_PAYS.xlsx**. Spécifique à chaque pays, ce fichier dresse la liste des combinaisons de produits et unités valides d'après le questionnaire NSU. En plus, il donne pour chaque produit et chaque unité, respectivement, son nom et son code.
* **produitRaccourciCode.xlsx**. Général à l'ensemble des pays du projet de l'enquête harmonisée, ce fichier contient les produits arrêtés, leurs noms, leurs codes, et les raccourcis utilisés pour ces produits dans les masques NSU.

### Comment installer

D’abord, ouvrir lancer\_FusionnerBasesNSU.do dans l'éditeur de Stata ou un autre éditeur de texte.



Ensuite, modifier les paramètres suivants :

* **projDir**. Mettre le chemin du répertoire où est le projet de fusion des données. Dans ce répertoire doivent se trouver les sous- répertoires suivants : donneesBrutes, donneesFinales, donneesProvisoires, etiquettes,
* **pays**. Mettre le nom de votre pays en bien respectant les valeurs admissibles indiquées en commentaires du programme.

Puis, sauvegarder lancer\_FusionnerBasesNSU.do afin de maintenir vos modifications.

### Comment utiliser

D’abord, lancer lancer\_FusionnerBasesNSU.do avec Stata par la méthode habituelle

Ensuite, vous retrouverez les bases suivantes dans le répertoire donneesFinales :

* **donnes\_rosterFixes.dta.** Contient toutes les informations concernant les unités énumérées à l’avance par pour chaque pays.
* **donnes\_rosterAutre1.dta.** Contient toutes les informations pour les unités connues, mais pas typiquement associées avec un produit donné.
* **donnes\_rosterAutre2.dta.** Contient les informations sur les unités nouvellement découverte sur le terrain lors de l’enquête NSU.
* **questionnaireMarche.dta.** Contient toutes les informations au niveau du roster.

Pour certaines enquêtes NSU, il existe un questionnaire de marché qui capte la localisation détaillée du marché de sorte à ne pas devoir capter ces mêmes informations dans chacun des 11 questionnaire couvrant un groupe de produits.

Pour le moment, le programme de fusion ne prend pas ce questionnaire en compte dans l’opération de fusion. Et pour utiliser le programme, il faut faire deux changements—un avant de lancer le programme, et un après l’exécution du programme :

* Avant de lancer le programme. Identifier le fichier zip (et éventuellement les bases) associé à cet instrument de collecte et le mettre en dehors du répertoire donneesBrutes
* Après exécution du programme. Faire des fusions manuelles entre les bases finales et la base de cet instrument. Pour ce faire, il faut identifier les variables d’identification de marché qui sont communes à une base finale et ce questionnaire, et fusionner les bases par ces variables. Voici un exemple avec des extraits fictifs de syntaxe en Stata :

\* identifier les variables communes entre la base roster fixe et la base marché

use “`donneesFinales’/ donnes\_rosterFixes.dta”, clear

d s00q\*

use “`donneesProvisoires’/detailsDuMarche.dta”, clear

d s00q\*

\* fusionner les bases avec les variables communes, et retenir les variables manquantes

use “`donneesFinales’/ donnes\_rosterFixes.dta”, clear

merge m:1 s00q01 s00q02 s00q03 s00q06 s00q07 using “`donneesProvisoires’/detailsDuMarche.dta”, keepusing(s00q04 s00q05)

\* sauvegarder le résulat

save “`donneesFinales’/ donnes\_rosterFixes\_NOUVEAU.dta”, replace

### Comment dépanner

Voir les consignes générales

En plus de cela, le programme crée un logs d’erreur, qui sont sauvegardés dans le répertoire « log »

* **fusionner\_errorLog.txt.** Capte des erreurs reliées à un format inattendu pour le nom des variables. Le programme de fusion s’attend à retrouver des variables dont le nom suit des formats bien clairs. Ce log capte des noms de variables qui s’écartent du schéma attendu.
* **fusionner\_typeVariable.txt.** Capte des problèmes divers dans les opérations du programme—notamment des problèmes dans le type de données pour certaines variables clé.

Ces fichiers informations servent à dépanner d’éventuels problèmes

## (Re)Classer les images (lancer\_classerNsuImages.do)

### Fonctionnement

Reclasse les images dans un nouveau système de répertoire de sorte à rendre le classement plus exploitable pour la séléction d'une image représentative d'un produit-unité.

En particulier, le programme :

* Vérifie le paramétrage
* Décomprime les archives d'images
* Crée un répertoire pour chaque produit
* Crée au niveau de chaque unité des répertoires d'unités
* Classe les images dans ce nouveau système de répertoires
* Modifie le nom de l'image afin de faire le lien entre l'image et le marché dans lequel elle a été prise.

### Description des composants

Composé des fichiers suivants :

* **lancer\_classerNsuImages.do**. Capte les paramètres à spécifier par l'utilisateur. Également, comme son nom l'indique, il lance le coeur du programme.
* **classerNsuImages\_v1\_INS.do**. Contient les instructions pour reclasser les images dans un nouveau système de répertoires
* **raccourcissement.doi**. Contient une syntaxe pour raccourcir ou autrement changer les noms de produits (e.g. enlever des parenthèses, mettre un nom plus compacte, etc.)
* **produits.dta**. Contient le nom, code, et raccourci répertorié pour le pays. Créé par le programme lancer\_FusionnerBasesNSU.do et hébergé dans le dossier de ressources.
* **unites.dta**. Contient le nom et code de chaque unité répertorié pour le pays. Créé lors du lancement de ce programme à partir de fichiers créés par le programme lancer\_FusionnerBases.do

### Comment installer

D’abord, ouvrir lancer\_classerNsuImages.do dans l'éditeur de Stata ou un autre éditeur de texte.

Ensuite, modifier les paramètres suivants :

* **projDir**. Mettre le chemin du répertoire où est le projet de fusion des données. Dans ce répertoire doivent se trouver les sous- répertoires suivants : donneesFinales, etiquettes, imagesBrutes, imagesClassees.
* **pays**. Mettre le nom de votre pays en bien respectant les valeurs admissibles indiquées en commentaires du programme.

Puis, sauvegarder lancer\_FusionnerBasesNSU.do

### Comment utiliser

Ne peut être utilisé qu'après avoir utilisé avec succès le programme lancer\_FusionnerBases.do, comme lancer\_classerNsuImages.do prend comme input certains outputs de lancer\_FusionnerBasesNSU.do

Lancer le programme lancer\_classerNsuImages.do avec Stata par la méthode habituelle

Voir les logs d'erreur, qui commencent par classer\_, dans le répertoire logs pour une liste des images que le programme n'a pas pu classer.

### Comment dépanner

Voir les consignes générales

En plus de cela, le programme crée un logs d’erreur, qui sont sauvegardés dans le répertoire « log »

* **classer\_logErreur.txt.** Capte de divers problèmes dans de déroulement du programme :
	+ Chemin de fichier trop long. Dépasse le nombre de caractères indiqués dans le programme. Le plafond absolu chez Windows est 255 caractères.
	+ Problème dans l’analyse du fichier. Le classement des fichiers repose sur le reconnaissance du produit à partir du nom de l’image. Si cette opération se solde en échec, le problème est documenté ici.
	+ Produit avec ce nom court n’existe pas dans la ressource. L’analyse du nom des images repose sur les noms courts de produit classés dans une ressource qui est une combinaison de deux fichiers Excel, un propre au pays et un général à l’ensemble des pays de l’UEMOA. Si le nom court du produit n’est pas retrouvé, on documente le problème.
	+ Unité avec ce code n’existe pas dans la ressource L’analyse du nom des images repose également sur le code des unités dans une ressource. Si le code de l’unité qui figure dans le nom de l’image n’est pas valide selon ce fichier ressource, on documente le problème.
	+ Image associé à une observation qui n’existe pas dans la base finale. Le nom des répertoires d’exportation correspondent au interview\_\_id dont l’image est issue. Si une observation avec une telle interview\_\_id n’existe pas dans les bases finale, le problème est documenté.
* **cheminTropLong.txt.** Un avertissement plus qu’une erreur, ce fichier compile les fichiers et répertoire de ce projet dont le chemin est assez long, mais pas trop long pour être utilisé.
* **nomsVides.txt.** Le fichier raccourcissement.doi dicte les changements ad hoc à appliquer au nom des produits. Si un produit n’a plus de nom après cette opération, le problème est documenté dans ce fichier.

**pasClasse.txt.** L’objectif de ce programme est de classer toutes les images de l’enquête dans un nouveau système de répertoires. Si cette opération ne marche pas, pour une raison ou une autre, le problème est documenté ici. Avec cette information, l’utilisateur peut reclasser les images manuellement, au besoin. Aussi, on peut voir les images à problème et ainsi constater des problèmes dans le programme.Ces fichiers informations servent à dépanner d’éventuels problèmes