|  |  |  |
| --- | --- | --- |
| logo_ec_17_colors_300dpi | EUROPEAN COMMISSION  EUROSTAT  Directorate F: Social statistics  **Unit F-4: Income and living conditions; Quality of life** |  |

|  |
| --- |
| **EU-SILC Validation Software**  **INSTALLATION GUIDE – EG and SAS 9.4 Users and R-studio users** |

Contents

[SILC VALIDATION INSTALLATION GUIDE 3](#_Toc212190084)

[Assumptions & Notations 3](#_Toc212190085)

[Directory Structure 4](#_Toc212190086)

[Directories Configuration 4](#_Toc212190087)

[Software Configuration - SAS Enterprise Guide 5](#_Toc212190088)

[SETENV-XML-EG.sas configuration 5](#_Toc212190089)

[SAS Application Server 5](#_Toc212190090)

[Software Configuration - SAS 9.4 Standalone (SA) Platforms 7](#_Toc212190091)

[Prerequisites – SAS Version Check 7](#_Toc212190092)

[SETENV-XML-SA.sas configuration 7](#_Toc212190093)

[Software Configuration – R-Studio Platforms 8](#_Toc212190094)

[Prerequisites – R-studio Version Check 8](#_Toc212190095)

[DATACHECKSQL.Rmd configuration 8](#_Toc212190096)

# SILC VALIDATION INSTALLATION GUIDE

## Assumptions & Notations

In this guide, we adopt the following assumptions/notations (the prefix “&” denotes an environment variable, which may differ according to the installation node, i.e.: to the NSI):

* **&eusilc** = **ROOT DIRECTORY** of your SILC installation (i.e.: C:\ or C:\SILC); this is the target directory of your installation process

THE CONTENT OF THE **INSTALLATION PACKAGE** MUST BE COPIED INTO THE root DIRECTORY[[1]](#footnote-1)

* **&cc** = country to be processed (e.g.: AT = Austria)
* **&ss** = transmission type, from 2021 onwards this parameter is set to R permanently
* **&F** = SILC File
  + D
  + R
  + H
  + P
* **&FVER** = SILC File Version
* **&yyyy**, **&yy** = survey year, respectively expressed as a 4-digit (i.e.: 2021) and a 2-digit figure (i.e.: 21)
* **&eusilc/main/&cc/csv** = inbound directory 🡪 the SILC data files to be processed are stored here and are expected to comply with the following naming rules:
  + SILC\_R**&F**\_A\_**&cc**\_**&yyyy**\_0000\_V**&FVER**.csv

Note: the file version here (&FVER) is padded with zeros accounting for 4 digits in total; for instance, a SILC file of version 1 is expressed as follows: \*V0001.csv

ALL FOUR SILC DATA files to be processed must have the same version number !

* **&eusilc/main/&cc/out** = output directory 🡪 at the end of the validation run, such directory encompasses the reports, either in printable – pdf or docx– or in spreadsheet form – csv
* **&eusilc/main/&cc/stage** = intermediate (“staging”) raw data sets directory 🡪 contains the data sets created as a result of the SILC data files elaboration
* **&eusilc/main/&cc/&ss&yy** = definitive raw data sets directory 🡪 contains the final data sets created as a result of the SILC data files elaboration; having output data sets in this directory is the precondition for the subsequent steps of the statistical chain, such as the BDB/IDB upload process and the Estimation Phase
* **&eusilc/XML\_CONFIG/XML** = XML configuration files directory; such files provide the configurable SILC validation rules
* **&eusilc/ XML\_CONFIG/MAP** = XML mapper directory; such files instruct SAS/EG about how to read the XML configuration files above mentioned
* **&eusilc/ XML\_CONFIG/XSD** = XSD files directory; the XSD files comes handy to edit the XML configuration files (if needed) and to check that they are structurally and semantically valid

## Directory Structure

The following screenshot shows the content of the SW installation bundle, **eusilc-V5.0.7z**. Its content must be transferred as-is to your hard disk so that its “eusilc-V5.0” directory be referenced by the **&eusilc** macro variable defined so far. The “eusilc-V5.0” directory shall then be renamed according to the macro variable **&eusilc**.

A group of yellow folders

AI-generated content may be incorrect.

It is worth remarking that the folder “**xx**” and its subfolders are empty. Said directory in fact is not meant to contain any SW component but just input (SILC data files to be validated) and output data (validation reports generated by the SW). The section “Directories configuration” below explains how to deal with it.

## Directories Configuration

In order to be able to execute the validation program, the fictitious country code “**xx**” should be renamed with “**&cc**”, which corresponds to your country two-digit code in lowercase letters.

# Software Configuration - SAS Enterprise Guide

The validation program is implemented as a SAS/EG (Enterprise Guide) project which in turn consists in the **DATACHECKXML.egp** file.

This can be run provided that a minimal configuration is accomplished. Read the next paragraph to these regards.

### SETENV-XML-EG.sas configuration

Once that the directory structure has been renamed, the context variable **&eusilc** must be set up within the configuration module, namely the following SAS module located in the **&eusilc/5.3\_Validation/pgm** directory:

* **SETENV-XML-EG.sas**

In order to provide the set up for said variable (“eusilc”), the following directive must be included on top of the SAS program in question, replacing \_\_\_PATH\_\_\_ with the actual location of the SILC ROOT DIRECTORY on your server:

* **%LET eusilc=\_\_\_PATH\_\_\_;**

The following illustrative directive refers to an installation environment having **C:\SILC** as the root directory:

* **%LET eusilc=C:\SILC;**

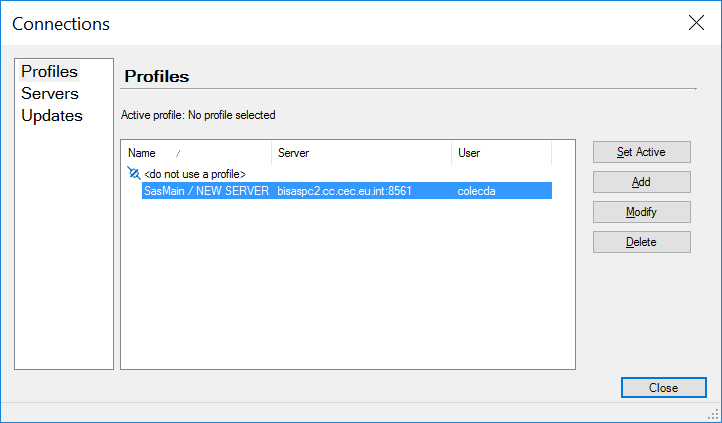
The last variable to set up is that aimed to specify whether the slash or the backslash is to be used as a directory separator:

* **%LET %LET \_dirsp\_=\;**
* **%LET %LET \_dirsp\_=/;**

The former setup applies in case of a Windows file system, whereas the latter applies in case of a UNIX file system.

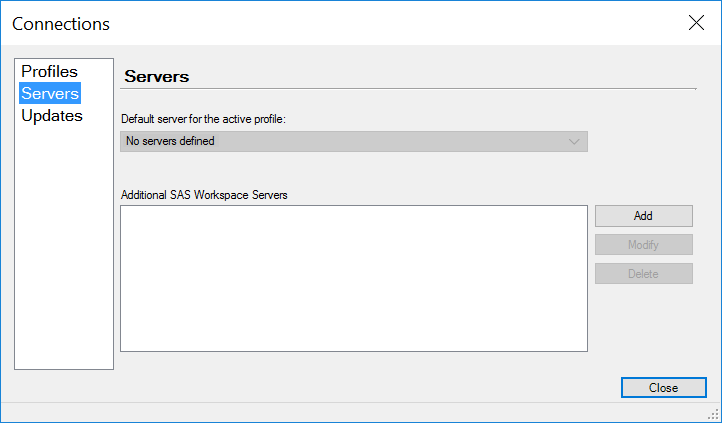
### SAS Application Server

Eurostat has developed the default SAS Application Server set up in the DATACHECKXML.egp. The user is supposed to overwrite this setup by selecting the proper SAS Application Server available in his execution environment. This operation can be accomplished through the menu **Tools** 🡪 **Connections** of SAS/EG which makes possible to choose the user’s SAS profile, as shown in the screenshot below:



IT arrangements may vary across NSIs. Please contact your IT department in case you do not have a SAS profile but you need your Organisation’s SAS Application Servers parameters.

These must be selected from the dropdown list shown below:



# Software Configuration - SAS 9.4 Standalone (SA) Platforms

### Prerequisites – SAS Version Check

The validation program is implemented through a collection of SAS programs (\*.sas files) which run on all SAS standalone platforms which should not be older than SAS 9.4. SAS older versions, such as SAS 9.3, 9.2 or SAS 9.1, are not compatible with the SILC Validation SW.

### SETENV-XML-SA.sas configuration

Once that the directory structure has been renamed, the context variable **&eusilc** must be set up within the configuration module, namely the following SAS module located in the **&eusilc/5.3\_Validation/pgm** directory:

* **SETENV-XML-SA.sas**

In order to provide the set up for said variable (“eusilc”), the following directive must be included on top of the SAS program in question, replacing \_\_\_PATH\_\_\_ with the actual location of the SILC ROOT DIRECTORY on your server:

* **%LET eusilc=\_\_\_PATH\_\_\_;**

The following illustrative directive refers to an installation environment having **C:\SILC** as the root directory:

* **%LET eusilc=C:\SILC;**

The last variable to set up is that aimed to specify whether the slash or the backslash is to be used as a directory separator:

* **%LET %LET \_dirsp\_=\;**
* **%LET %LET \_dirsp\_=/;**

The former setup applies in case of a Windows file system, whereas the latter applies in case of a UNIX file system.

# Software Configuration – R-Studio Platforms

### Prerequisites – R-studio Version Check

The validation program is implemented through a collection of R programs (\*.r files) which run on R-Studio platforms which should not be older than R-Studio 2023.03.0.

The following packages are required to run the R prevalidation: **sqldf, DBI, dplyr, DT, htmltools, tinytex, knitr**.

### DATACHECKSQL.Rmd configuration

Once that the directory structure has been renamed, the context variable **EUSILC** must be set up within the configuration module, namely the following R module located in the **EUSILC/5.3\_Validation/R** directory:

To provide the set up for said variable (“EUSILC”), the following directive must be included on top of the R program in question, replacing \_\_\_PATH\_\_\_ with the actual location of the SILC ROOT DIRECTORY on your server:

* **WorkingDrive: "\_\_\_PATH\_\_\_"**

The following illustrative directive refers to an installation environment having **C:/SILC** as the root directory:

* **WorkingDrive: "C:/SILC"**

1. The subdirectories’ structure must be obviously preserved. [↑](#footnote-ref-1)