|  |  |  |
| --- | --- | --- |
| 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**  **USER GUIDE –Struval users** |

Table of contents

[Introduction 3](#_Toc181286756)

[1. Preliminary information 4](#_Toc181286757)

[1.1 Methodological references 4](#_Toc181286758)

[1.2 SILC operation years 4](#_Toc181286759)

[1.3 SILC Data Files 4](#_Toc181286760)

[1.4 Useful operational notations 4](#_Toc181286761)

[2 Running the pre-validation in STRUVAL 5](#_Toc181286762)

[2.1 EDAMIS 5](#_Toc181286763)

[2.2 User-guide 5](#_Toc181286764)

[3 Running the pre-validation in R-Studio 6](#_Toc181286765)

[3.1 The results in R-Studio 6](#_Toc181286766)

[3.2 Results datasets 7](#_Toc181286767)

[ANNEX 1: The R-Studio pre-validation programs structure 10](#_Toc181286768)

[ANNEX 2: comparison of checks implemented in each process 11](#_Toc181286769)

# Introduction

Following the entry into force of IESS [Regulation (EU) 2019/1700 establishing a common framework for European statistics relating to persons and households, based on data at individual level collected from samples](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1700&from=EN), in 2021 Eurostat has re-engineered the internal IT tool to validate the data of the European Union Survey on Income and living conditions (EU-SILC). This validation software is based on SAS technologies (SAS enterprise guide or SAS foundation).

However, as many NSIs do not or cannot use SAS, an alternative process, albeit less complete, has been developed to enable them to pre-validate their SILC data.

This process is **two-stepped**:

* A first part of the pre-validation checks is implemented in STRUVAL
* The remaining part of the pre-validation checks is implemented in R programs

This User Guide aims at providing EU-SILC data validators (internal Eurostat’s EU-SILC production team members and colleagues from the NSIs) with practical, operational information on the first step of the EU-SILC alternative pre-validation with STRUVAL.

The document presents the following outline:

1. Information on methodological references, operation years, SILC data files and notation to be followed to properly execute the pre-validation.
2. Instruction on how to run the 1st step of pre-validation in STRUVAL.
3. Annex 1 reports a comparison of the checks performed in the SAS programs with the checks performed when using R-Studio and STRUVAL.

Eurostat invites colleagues from NSIs to contact the EU-SILC production team ([ESTAT-ILC@ec.europa.eu](mailto:ESTAT-ILC@ec.europa.eu)) for any information needed on the presented pre-validation process.

# Preliminary information

# **Methodological references**

For the User’s information, the pre-validation process is based on a number of validation rules checking the compliance of SILC variables, values and flags reported by data providers in the SILC files with the reference variables/flags and relative modalities contained in an XML file. Such file is annually configured according to the [EU-SILC Methodological guidelines (DOC065)](https://circabc.europa.eu/ui/group/853b48e6-a00f-4d22-87db-c40bafd0161d/library/334d943f-6f71-4f4b-9c7e-a6767a3fe164?p=1&n=-1&sort=name_DESC).

# **SILC operation years**

As SILC provides cross-sectional and longitudinal information[[1]](#footnote-1), the process is configured to refer to the proper validation rules for each operation year.

In other terms, if a country in December 2025 provides a 4-year panel SILC data, its data file will report information for the last 4 years (2022, 2023, 2024 and 2025 – or back to 2020 in case of 6 year panel).

# **SILC Data Files**

NSIs send EU-SILC data in four separate files:

* Household register (D)
* Personal register (R)
* Household data (H)
* Personal data (P)

The household register file (D) contains every household selected, including those where the address could not be contacted and households that could not be interviewed.

In the other files, records associated with a household will exist only if the household has been contacted and has a completed household interview in the household data file (H) and at least one member has complete data in the personal data file (P). This member must be the selected respondent if this mode of selection is used.

The personal register file (R) contains a record for every person currently living in the household or temporarily absent. As regards the longitudinal component, it must also contain a record for every person registered in the previous year’s R-file.

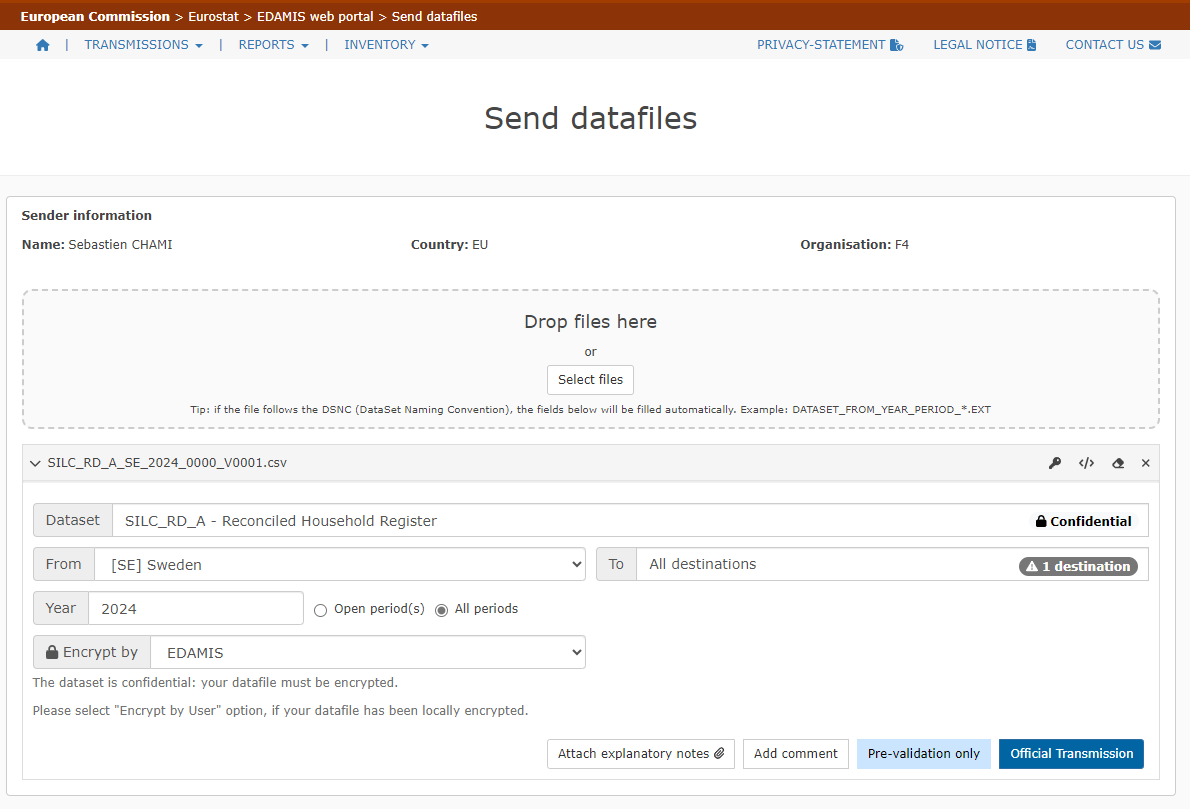
Finally, the personal data file (P) contains a record for every eligible person for whom the information could be completed from interview and/or registers.

# Running the pre-validation in STRUVAL

# **EDAMIS**

The STRUVAL pre-validation checks are implemented on the EDAMIS platform which is already used for official EU-SILC data transmissions: <https://webgate.ec.europa.eu/edamis4/>

**Figure 1: EU-SILC STRUVAL prevalidation on EDAMIS**



# **User-guide**

A complete and comprehensive user guide for prevalidation on Edamis is available here: <https://cros.ec.europa.eu/book-page/edamis-pre-validation-user-guide>

and the list of error codes: [STRUVAL error codes and messages | CROS](https://wayback.archive-it.org/12090/20221109161355/https:/ec.europa.eu/eurostat/cros/content/struval-error-codes-and-messages_en)

# ANNEX 1: comparison of checks implemented in each process

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Check** | | **SAS** | **STRUVAL** | **R-studio** |
| Expected variables | | **✓** | **✓** | **X** |
| Unexpected variables | | **✓** | **✓** | **X** |
| S  V  A  L | IDs uniqueness | **✓** | **✓** | **X** |
| Values | **✓** | **✓** | **X** |
| Flags | **✓** | **✓** | **X** |
| Imputation factors | **✓** | **!** | **X** |
| Value vs flag | **✓** | **X** | **✓** |
| Routing conditions (-2 and -4 flags) | **✓** | **X** | **✓** |
| Non-selected respondents (-3 flags) | **✓** | **X** | **✓** |
| Discontinued and created variables (-7 flags) | **✓** | **X** | **X** |
| Household grid | **✓** | **X** | **✓** |
| Logical checks | | **✓** | **X** | **✓** |
| Sample-size, Weights, Outliers, Counting, Comparison, Missing rates | | **✓** | **X** | **X** |
| Summary | | **✓** | **!** | **!** |

Legend

|  |
| --- |
| **✓ : checks fully implemented** |
| **! : checks partially implemented** |
| **X : checks not implemented** |

1. Cross-sectional data refer to a given time or a certain time period with variables on income, poverty, social exclusion and other living conditions. Longitudinal data refer to individual/household changes over time, observed periodically over a four-year period (or more years if a longer duration panel is used). [↑](#footnote-ref-1)