TWINNING CONTRACT

2023/447-234

EU for Further Development of Statistics System in BiH

MISSION REPORT

Activity 4A: Calculation of SILC indicators (SAS)
Component 4: Statistics on Income and Living Conditions

Mission carried out by

Veli-Matti Törmälehto, Statistics Finland
Marie Reijo, Statistics Finland

04 – 06 March 2025

Version: Final

European Union Member State Expert Contact Information

Veli-Matti Törmälehto
Statistics Finland
 veli-matti.tormalehto@stat.fi

Marie Reijo
Statistics Finland
marie.reijo@stat.fi

Table of contents

List of Abbreviations

BC	Beneficiary Country (Bosnia and Herzegovina)
BHAS	Agency for Statistics of Bosnia and Herzegovina
BiH	Bosnia and Herzegovina
CBBH	Central Bank of Bosnia and Herzegovina
EU	European Union
EUD	European Union Delegation to Bosnia and Herzegovina
FIS	Institute for Statistics of the Federation of Bosnia and Herzegovina
LA	Language Assistant
MS	EU Member State
RSIS	Institute for Statistics of Republika Srpska
RTA	Resident Twinning Advisor
RTAA	Resident Twinning Advisor Assistant
ToR	Terms of Reference

Executive Summary

The focus of the mission was on calculation of SILC indicators. Statistics Finland presented their approach in calculating the indicators as well as the Eurostat indicators production tool programmed in SAS. During the mission, the tool was also used as a hand-on exercise to compute indicators for BiH 2022 SILC data. It was concluded that it is possible to run the Eurostat indicators production chain in SAS by the BiH staff. Statistics Finland also presented their SILC production database, and gave an introductory presentation on quality reporting requirements, in particular the domain-specific requirements of EU-SILC. These should be further discussed in the coming missions. Following active discussions with the BiH staff, the consultants proposed five recommendations related to calculation of indicators, data management, and software tools. A key recommendation is to prepare a migration plan for moving SILC production from SPSS to R, with the aim to complete the migration and to create the new required tools within 1 to 3 years.

1. General comments

This mission report was prepared within the EU Twinning Project "Further Development of the Statistics System in Bosnia and Herzegovina". It was the first mission devoted to the Statistics on Income and Living Conditions (SILC) Component of the Project.

The purposes of the mission were:

- Presentation of calculation of SILC indicators in Finland
- Introduction to quality reporting
- Mission Report
- Preparation of ToR for next activity

The consultants would like to express their thanks to all officials and individuals from Bosnia and Herzegovina met for the kind support and valuable information, and which highly facilitated the work of the consultants.

This views and observations stated in this report are those of the consultants and do not necessarily correspond to the views of EU, BHAS, FIS, RSIS, CBBH, Statistics Denmark, Statistics Finland, Statistical Office of the Republic of Slovenia or The Italian National Institute of Statistics.

2. Assessment and results

In the beginning of the mission, the current state of play regarding SILC (Survey on Income and Living Conditions) was reviewed by colleagues from BiH. Three fieldworks of SILC have been successfully completed in 2022-2024. Cross-sectional databases of 2022 and 2023 have been processed, while 2024 data are still in raw stage. Panel length is four years, and CAPI (Computer Assisted Personal Interviews, Blaise software) is the main collection mode. The main problem is a lack of proper tools to process the data. SPSS has been used by BiH in the processing, but it seems evident that it is not ideal for SILC data processing and management, as there are no Eurostat SPSS tools available for SILC production (tools are available only in SAS, partly also R).

EU FOR FURTHER DEVELOPMENT OF STATISTICS SYSTEM IN BIH

4

Funded by the European Union · STATISTICS DENMARK · Istat · REPUBLIC OF SLOVENIA STATISTICAL OFFICE · Statistics Finland

The focus of the mission was on calculation of SILC indicators. Current Eurostat indicators code in SAS have been adapted to SPSS by BiH, but this takes quite some effort and the results have not been the same that Eurostat produces.

Statistics Finland short-term experts presented their approach to calculation of indicators at Statistics Finland. Statistics Finland focuses on a subset of indicators, which have precision requirements in the EU-SILC legislation or which need to be validated before Eurostat publishing, such as at risk of poverty (AROP), persistent at risk of poverty etc. It was clarified that while Eurostat indicators code in SAS is available from CIRCA, some adaptations have been necessary, for instance to be able to compute standard errors.

Regarding the software, Statistics Finland uses now SAS to process and analyze data stored in Microsoft SQL Server database. However, in the coming years Statistics Finland will migrate all statistical productions to new platforms and technologies, and Python/R /SQL will replace SAS as the main software tools. Similar changes are taking place in other statistical organizations, including Eurostat, and there is a general movement towards open-source software solutions. It was however noted that SQL as a scripting language is common to all platforms/software environments.

Statistics Finland short-term experts continued by reviewing the main steps of the Eurostat SAS indicators program code (main script to run the validations). It was noted that a key step is the setup of the input data, which the code then transforms to base data for calculation of the indicators. The output consists of indicators database and RTF-format output in Word, which shows indicators up to four years and their changes. The report can be used to detect large changes or anomalies in the indicators. It was also noted that the current Eurostat SAS code largely builds on sql queries; moreover, the available R validation codes from Eurostat also rely on sql queries.

As a hand-on exercise, the 2022 SILC data of Bosnia and Herzegovina was used to test the Eurostat indicators SAS code. This was feasible with a trial version of SAS available on one of the BiH laptops. The files in SPSS were first converted to SAS and copied to a country-specific folder. These were then successfully converted to the input files using the Eurostat code, which BiH was able to download from CIRCA. After this, the SAS code links the different files, and at this point some modifications to the code were necessary. These consisted essentially just renaming of two variables. The data linkage was then successfully completed, and a subsequent step which adds Eurostat variables to the data also worked without problems. Some indicators were successfully computed by BiH. It was concluded that it is possible to run the Eurostat indicators production chain in SAS by the BiH staff as long as they have an active SAS license. However, some of the metadata files included in the Eurostat package do not include information for BiH (e.g. purchasing power parities and population sizes), and these should be added as they are required for some of the indicators.

Following a request from the BiH staff, a presentation on basic features of SILC production database at Statistics Finland was added to the agenda. Data are organized in relational MS SQL Server database in Finland, and a new database is set up for each year. Relational database (as opposed to a file system) seems necessary in Finland, given the complex structure of the SILC data with both cross-sectional and longitudinal comments and household and individual data. The collected data (Blaise interviews, administrative sources) are extracted and loaded to several primary database tables, with only minor transformations before the loading. Data processing is implemented within the database (editing, imputing, modelling, deriving aggregates etc.). SILC tables are then derived from the primary tables (four core tables and separate tables for modules).

The Finnish EU-SILC database already follows the so-called reconciled EU-SILC format regarding the final SILC data. There are no separate longitudinal tables or files. Longitudinal data are integrated within the cross-sectional data files, and these include longitudinal variables which can be used to link the panels across the SILC operations and to compute the longitudinal indicators. It was noted that Finland also follows a four-year panel in SILC.

Statistics Finland short-term experts gave an introductory presentation on quality reporting requirements, in particular the domain-specific requirements of EU-SILC. The standard form of the quality reporting by the producers to Eurostat is SIMS (Single Integrated Metadata Structure). Eurostat then disseminates both national and comparative EU-SILC quality reports. Eurostat has SAS tools to compute EU-SILC quality measures and indicators, which can be installed parallel to validation and indicator production chains. National quality report is transmitted via a Web portal.

At the moment, BiH has not yet implemented the SILC quality reporting, but they have done so for other statistical domains and are well aware of the overall requirements for Eurostat quality reporting. SAS program workflow for SILC quality reporting may be a useful topic to be discussed in the coming missions, similarly to indicators and validations. Computation of standard errors may require specific tools, which take into account the sampling design and weighting in BiH and is probably the most challenging part of quality reporting because some of the SILC indicators are quite complex.

3. Conclusions and recommendations

Recommendation 1: Complete implementation of the Eurostat's indicator production chain in SAS, for all years that Bosnia and Herzegovina has SILC data.

Recommendation 2: Create a plan for organizing the SILC data, either into a relational database or in a standardized and improved version of the current file system. The aim should be the reconciled SILC format (Eurostat standard), with cross-sectional and longitudinal data integrated into the same files for each SILC operation.

Recommendation 3. Prepare a migration plan for moving SILC production from SPSS to R, with the aim to complete the migration and to create the new required tools within 1 to 3 years. R is recommended because it is widely used, already supported by Eurostat for data validation, and it is more suitable for survey data than some alternatives (e.g. Python). The migration plan should, however, take into account developments in Eurostat as they are also in transition phase at the moment (moving away from SAS 9.4). Migration will require significant amount of work from IT experts and subject-matter experts, and these resource requirements should be explicitly accounted for in the migration plan.

Recommendation 4. Extend the current SAS license to an additional year, and consider adding one license for an IT expert. This is needed as an intermediate step (in 2025 and 2026) to benchmark the results of migration (to R) with the results produced using the Eurostat's SAS tools.

Recommendation 5: In the indicators, consider focusing on the most relevant subset of indicators, i.e. those that have precision requirements in the EU-SILC legislation (AROP, AROPE, incl. breakdown by NUTS2). A method to calculate standard errors should also be developed (in R, making use of already available packages).

4. What to do before the next mission

Action	Deadline	Responsible person
Establish data flow and folder structure for calculation of SILC indicators in a standardised way (following Eurostat's guidelines in the indicator package)	Before the next mission	BHAS (in collaboration with FIS and RSIS).
Calculate all cross-sectional indicators for all available years 2022-	Before the next mission	BHAS (in collaboration with FIS and RSIS).
A preliminary design for the revised SILC database, incl. longitudinal component	Before the next mission	BHAS (in collaboration with FIS and RSIS).

Annex 1. Terms of Reference for the current mission

<div style="border:1px solid black;">

Terms of Reference

EU Twinning Project BA 19 IPA ST 01 23

Component 4: Statistics on Income and Living Conditions

Timing:
04-06 March 2025

Venue: Agency for Statistics of BiH, Pehlivanusa 3, Sarajevo
Bosnia and Herzegovina

</div>

Activity 4A: Calculation of SILC indicators (SAS)

Initially, the Project was supposed to assist on utilizing administrative data (AD) sources for the compilation of indicators on wages and salaries. However, since this topic is also on the agenda in the AD project, the BC has requested to change the content and, consequently, the results in the Twinning Project. The new requested subject is EU SILC (Statistics/Survey on Income and Living Conditions) focusing only on the survey part (Living Conditions), as the indicators for income will be covered by the AD project.

The Project will provide expert assistance in the processing of EU SILC research (production of quality reports, assistance in the preparation of longitudinal weights, assistance in the calculation of basic indicators).

1. **Mandatory results**
Preparation of quality report.
Creation of longitudinal weights.
Calculation of SILC indicators.

2. **Purpose of the activity:**

- Presentation of calculation of SILC indicators in Finland
- Introduction to quality reporting
- Mission Report
- Preparation of ToR for next activity

EU FOR FURTHER DEVELOPMENT OF STATISTICS SYSTEM IN BIH

8

Funded by the European Union | STATISTICS DENMARK | Istat | REPUBLIC OF SLOVENIA STATISTICAL OFFICE | Statistics Finland

3. Expected outputs of the activity:

- Presentation of calculation of SILC indicators in Finland delivered
- Introduction to quality reporting presented
- Mission Report produced
- ToR for next activity prepared

Annex 2. (input for the) Terms of Reference for the next mission

Activity 4B	Production of Longitudinal Dataset (SAS)
Budget section	III. Components Mandatory Results – Component 4
Subject	• Production of longitudinal dataset (SAS) and validations (to be confirmed) • Quality reporting (further discussions) • Evaluation of the actions required after the 1st mission • Mission Report • Preparation of ToR for next activity
Methods	Onsite, STE mission in BiH
Resources	MS: Marie Reijo and Sonja Bärlund (to be confirmed) FI, RTA, RTAA and RTALA BC: Relevant staff expected to be involved in component 4 Venue: Banja Luka (to be confirmed)
Duration	2 MS STEs x 3 working days
Output	• Production of longitudinal dataset (SAS) presented • Mission Report produced • ToR for next activity prepared
Time schedule	Second week of September 2025 (to be confirmed)

Annex 3. Persons met

BHAS:
- Radoslav Ćorović, Assistant Director in the Sector for Demography and Social Statistics
- Zlatan Hadžić, Expert Advisor for the Statistics on Income and Living Conditions (SILC)
- Vladimir Mijović, Head of Labour Market Department

FIS:
- Amela Vesković, Head of the Unit of the Living Conditions
- Mirza Omanić, Unit of the Living Conditions
- Hidajeta Čolović, Assistant Director

RSIS:
- Vanja Vilipić, Unit of the employment and social statistics
- Biljana Tešić, Unit of employment and social statistics
- Pero Kazanović, IT expert in RSIS

RTA Team:
- Niels Madsen, RTA
- Larisa Muslimović, RTA Assistant
- Senka Ahmetović-Palić, Language Assistant

Signatures

For the approval of the contents of this report, representatives from BHAS, FIS and RSIS as well as MS experts and the RTA sign here:
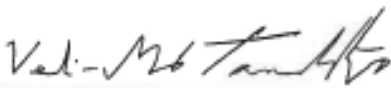

Zlatan Hadžić
_____
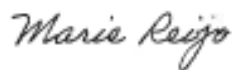*Component leader, BHAS*

_____
Component leader, FIS

Биљана Тешић
_____
Component leader, RSIS

_____
RTA

_____
MS Expert

Marie Reijo
_____
MS Expert

Annex 4: List of mission presentations

- Calculation of SILC Indicators, Marie Reijo
- Introduction to Quality Reporting, Marie Reijo
- Basic Features of SILC Production Database at Statistics Finland, Veli-Matti Törmälehto