



# What's new about Eurostat's European microdata?

8<sup>th</sup> European User Conference for EU-Microdata

16-17/03/2023

Hartmut Schrör, Emilio Di Meglio, Frank Espelage

# Contents

1. New and changed scientific use files
  - EU-LFS
    - Datasets and (sub)samples
    - Variables and transcoding
    - Anonymisation criteria
  - EU-SILC
  - CVTS
2. Remote access to European microdata (secure use files)

# EU-LFS: Datasets and changes with IESS

- Old SUFs up to 2021 release (and reference year 2020)
  - Quarterly datasets for the quarterly variables only
  - Yearly datasets
  - Modules
  - Household data (for countries using household subsamples)
- New SUFs from 2022 release (covering all reference years)
  - Quarterly datasets for the quarterly variables only
  - Yearly datasets, including all module and household data
    - => more complex, but all data in one file

# EU-LFS: Annual subsamples of full quarterly samples

- **Subsampling** requirements and constraints **from reference year 2021**:
  - Allowed for yearly, biennial, household information
  - Compulsory for regular modules / ad-hoc subjects
- (Sub)-samples for results at individual level:
  - quarterly  $\geq$  yearly  $\geq$  biennial  $\geq$  module
- For results at household level:
  - Minimum set of variables (subset of quarterly and yearly variables)
- Certain modules (e.g. migrants, reconciliation of work and family life): some household composition information to be provided

# EU-LFS: Annual (sub-)samples and weights

01	Quarterly [only in the quarterly dataset]	COEFFQ
02	Quarterly and yearly	COEFFY, (COEFFHH)
03	Quarterly, yearly and biennial	COEFFY, COEFF2Y, (COEFFHH)
04	Quarterly, yearly, biennial and module	COEFFY, COEFF2Y, COEFFMOD, (COEFFHH)
05	Quarterly and (originally selected) respondent forms part of household subsample	COEFFHH
06	Quarterly, yearly and (originally selected) respondent forms part of household subsample	COEFFY, COEFFH
07	Quarterly, yearly, biennial and (originally selected) respondent forms part of household subsample	COEFFY, COEFF2Y, COEFFHH
08	Quarterly, yearly, biennial, module and (originally selected) respondent forms part of household subsample	COEFFY, COEFF2Y, COEFFMOD, COEFFHH
09	Household - minimum set of variables (for additional household members)	COEFFHH
10	Household - restricted set of module background variables (for additional household members)	-

# EU-LFS: How to select subsamples in yearly datasets

- Yearly variables:  $\text{COEFFY} > 0$
- Biennial variables:  $\text{COEFF2Y} > 0$  (from 2021)
- Module:  $\text{COEFFMOD} > 0$
- Household:  $\text{COEFFHH} > 0$ 
  - $\text{COEFFHH}$  for individual information
  - $\text{COEFFHHAVG}$  for household level information

**Weights may be identical depending on the subsampling options**

# Contents

1. New and changed scientific use files
  - EU-LFS
    - Datasets and (sub)samples
    - **Variables and transcoding**
    - Anonymisation criteria
  - EU-SILC
  - CVTS
2. Remote access to European microdata (secure use files)

# EU-LFS: Transcoding of variables

- From reference year 2021 (new regulation), variables have been dropped / changed / added to the EU-LFS; some changed periodicity.
- Variables for years < 2021 have been transcoded accordingly to the extent possible to have consistent time series.
- Constraints to availability: household data from 2006, modules from 2005.
- Detailed documentation of transcoding

[EU-LFS User Guide](#)

[Transcoding tables](#)



# IESS: definitions, concepts, variables

- “Standardised key social variables” have been harmonised across the EU social surveys on individuals / households
- LFS variables: technical items (MODE, PROXY, DEGURBA, etc.), citizenship and migration, job characteristics, education, health.  
-> 33 of 116 core variables
- Details in [Commission Implementing Regulation \(EU\) 2019/2181](#) and related [implementing guidelines](#).

# 8-yearly variables and ad hoc subjects

2021: Labour market situation of migrants and their immediate descendants

2022: Job skills (ad hoc subject)

2023: Pensions and labour market participation

2024: Young people on the labour market

2025: Reconciliation between work and family life

2026: (ad hoc subject to be decided)

2027: Work organisation and working time arrangements

2028: Accidents at work and work-related health problems

# Contents

1. New and changed scientific use files
  - EU-LFS
    - Datasets and (sub)samples
    - Variables and transcoding
    - **Anonymisation criteria**
  - EU-SILC
  - CVTS
2. Remote access to European microdata (secure use files)

# EU-LFS anonymisation criteria

- Criteria revised for the 2022 release in order to
  - introduce detailed age information
  - adapt the criteria to the variables in the new regulation to the extent necessary
- AGE
  - Available in single years, with some country-specific constraints: age groups, top-coding, perturbation (within 5-year age groups)
  - AGE\_GRP (5-year age groups) provided in addition
- All detail is in the [EU-LFS User Guide](#), chapter 7

# Contents

1. New and changed scientific use files
  - EU-LFS
    - Datasets and (sub)samples
    - Variables and transcoding
    - Anonymisation criteria
  - **EU-SILC**
  - CVTS
2. Remote access to European microdata (secure use files)

# What is new in EU-SILC?

- Change in the legal basis (IESS regulation)
- Some variables are added, some are removed
- Annual **nucleus** variables (90%) and **modules** (10%) - every three years, every six years and ad-hoc subject / policy need modules, additional variables could be added as optional (e.g.: in 2020 and 2021 - > COVID -19 variables).
- Change in precision requirements
- Introduction of regional precision requirements

# EU-SILC Legal basis

- ✓ Legal basis has been revised for all social statistics and new legislative framework came into force in 2021

**Regulation (EU) 2019/1700 of the European Parliament and of the Council of 10 October 2019 establishing a common framework for European statistics relating to persons and households, based on data at individual level collected from samples [Legislation - Income and living conditions - Eurostat \(europa.eu\)](#)**

## Policy relevance

- Simplifying and integrating
- Reinforcing quality and comparability
- Standardization
- Efficiency
- Timeliness
- Precision
- Regional dimension



# EU-SILC content

Annual core (nucleus)	3-yearly Rolling module	6-yearly Rolling module	Ad-hoc policy subjects	Year
<ul style="list-style-type: none"> <li>- Income</li> <li>- Deprivation</li> <li>- Economic Activity</li> <li>- Demography</li> <li>- Education</li> <li>- Child care</li> <li>- Housing costs</li> <li>- Health</li> <li>- Quality of life</li> <li>- Miscellaneous</li> <li>- Technical variables</li> </ul>	Children		Living arrangements and conditions of children in separated and blended families	2021
	Health	Quality of life		2022
	Labour & Housing conditions	Intergenerational transmission of disadvantages & Housing difficulties	Energy poverty	2023
	Children	Access to services		2024
	Health		Energy and environment	2025
	Labour & Housing conditions	Over-indebtedness, consumption and wealth		2026



# Newly approved legislation for 2024 Module

## 6-year Rolling Module starting in 2024 on Access to Services

Document 32022R2498

**Commission Implementing Regulation (EU) 2022/2498 of 9 December 2022 specifying technical items of data sets of the sample survey in the income and living conditions domain on access to services pursuant to Regulation (EU) 2019/1700 of the European Parliament and of the Council (Text with EEA relevance)**

C/2022/8948

OJ L 325, 20.12.2022, p. 16–30 (BG, ES, CS, DA, DE, ET, EL, EN, FR, GA, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV)

● In force

ELI: [http://data.europa.eu/eli/reg\\_impl/2022/2498/oj](http://data.europa.eu/eli/reg_impl/2022/2498/oj)

⌵ Languages, formats and link to OJ

	BG	ES	CS	DA	DE	ET	EL	EN	FR	GA	HR	IT	LV	LT	HU	MT	NL	PL	PT	RO	SK	SL	FI	SV
HTML																								
PDF																								
Official Journal																								

To see if this document has been published in an e-OJ with legal value, click on the icon above (For OJs published before 1st July 2013, only the paper version has legal value).

### Contents:

- Use of services, including care services and services for independent living
- Affordability of services
- Unmet needs and reasons

# Changes in variables

- RB081 Age in completed years (top coded)
- RB082 Age in completed years at the time of the interview (top coded)
- PB240 Country of birth of mother (recoded as “LOC”, “EU”, “OTH”)
- PB230 Country of birth of father (recoded as “LOC”, “EU”, “OTH”)
- RB285 Duration of stay in the country of residence (5-year groups)
- PE041 Educational attainment level (top code at ISCED 500)
- PL111B Economic activity of the local unit (last job) recoded

# Changes in variables

- RG Household grid
- PB205 Partners living in the same household
- HB110 Household type
- PL271 Duration of the most recent unemployment spell
- PL016 Existence of previous employment experience
- HS022 Reduced utility costs
- HY121 Taxes paid on ownership of household main dwelling
- HI010 Change in the household income compared to previous year
- HI020 Reason for increase in income
- HI030 Reason for decrease in income
- HI040 Expectation of the household income in the next 12 months

# Calendar of UDB

- Currently 2 releases per year: spring and autumn
- With the entry into force of IESS regulation the new transmission deadline for NSIs is 28/2
- Work in progress to gradually move towards one UDB per year at June/July
- Probably achieved in 2025

# Updates in EU-SILC Methodology page

Overview - Income and living conditions

ec.europa.eu/eurostat/web/income-and-living-conditions/overview

europa.eu

Log in EN English Enter search term Search

Home Data News Publications About us Contact us Help

Home > Data > Statistical themes > Income and living conditions > Overview

## Income and living conditions – Overview


INTRODUCTION

### What information can I find here?

The European statistics on income, social inclusion and living conditions cover objective and subjective aspects in monetary and non-monetary terms for households and individuals. The following aspects are covered:

- income
- poverty
- social exclusion
- housing
- labour
- education
- health

The main source for the compilation of statistics on income, social inclusion and living conditions is the EU-Statistics on Income and Living Conditions (EU-SILC) instrument.



HIGHLIGHTS

#### Compare yourself to other Europeans

This visualization tool 'You in the EU' allows you to easily compare your own living conditions to other people in your country and in the EU.

#### Learn about at risk of poverty or social exclusion

With this bar chart you can find out if the share of people at risk of poverty or social exclusion in your country is higher or lower compared with others.

#### Discover what's behind EU-SILC

Have a look at this helpful online publication in Statistics Explained describing background information on EU SILC, such as concepts, data collection or quality.

DIRECT ACCESS TO...

- Database
- Modules
- Information on data
- Publications
- Statistics Explained
- Methodology
- Legislation

EXPLORE FURTHER

#### Focus on quality

Have a look at the EU and national quality reports with information on different related aspects.

# Contents

## 1. New and changed scientific use files

- EU-LFS
  - Datasets and (sub)samples
  - Variables and transcoding
  - Anonymisation criteria
- EU-SILC
- **CVTS**

## 2. Remote access to European microdata (secure use files)

# CVTS – Continuing Vocational Training

- “Lifelong learning” statistics consist of CVTS, AES, LFS
- Scientific use files on CVTS 6 (2020) are available
  - CVTS is an enterprise survey that takes place every 5 years
- 25 countries: EU-27 except IE, HR, NL, AT; NO and RS
- Coverage: enterprises in the business economy with 10 or more persons employed, net sample about 113,000 enterprises
- Topic: enterprises’ investment in the continuing vocational training of their staff
- Definition of CVT: education or training activities financed in total or at least partly by the enterprise

# CVTS – Continuing Vocational Training

- Data on
  - Provision of CVT courses and other forms of CVT (such as guided on-the-job training, job rotation, exchanges, ...)
  - CVT strategies
  - Details on CVT courses: participants, costs, hours spent, characteristics such as skills targeted, providers
  - Assessment of CVT activities
  - Factors limiting or reasons for not providing CVT
  - Some information on initial vocational training (IVT)
- References

[https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Continuing Vocational Training Survey \(CVTS\) methodology](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Continuing_Vocational_Training_Survey_(CVTS)_methodology)



# Contents

## 1. New and changed scientific use files

- EU-LFS
  - Datasets and (sub)samples
  - Variables and transcoding
  - Anonymisation criteria
- EU-SILC
- CVTS

## 2. Remote access to European microdata (secure use files)

# Remote access to European microdata

## Motivation:

- Traditionally: researcher access to **secure use files** at Eurostat premises (safe centre)
  - Difficult: travel to Luxembourg, limited time in safe centre, 'host' at Eurostat...
  - Impossible during pandemic/lockdown periods
- Aim: facilitate the use of confidential statistical data for scientific purposes
- European Statistical System Committee (ESSC) approved the idea of providing remote access to European microdata in 2021

# Remote access to European microdata

Legal basis:

- Commission Regulation (EU) No. 557/2013 on access to confidential data for scientific purposes defines general conditions of access to:
  - Scientific use files
  - Secure use files
- Access to secure use files is only allowed from/within (Art.7,8):
  - European Commission access facilities such as Eurostat's safe centre
  - **Other access facilities accredited by the European Commission**
- Research output checking is required for secure use files (Art.7)

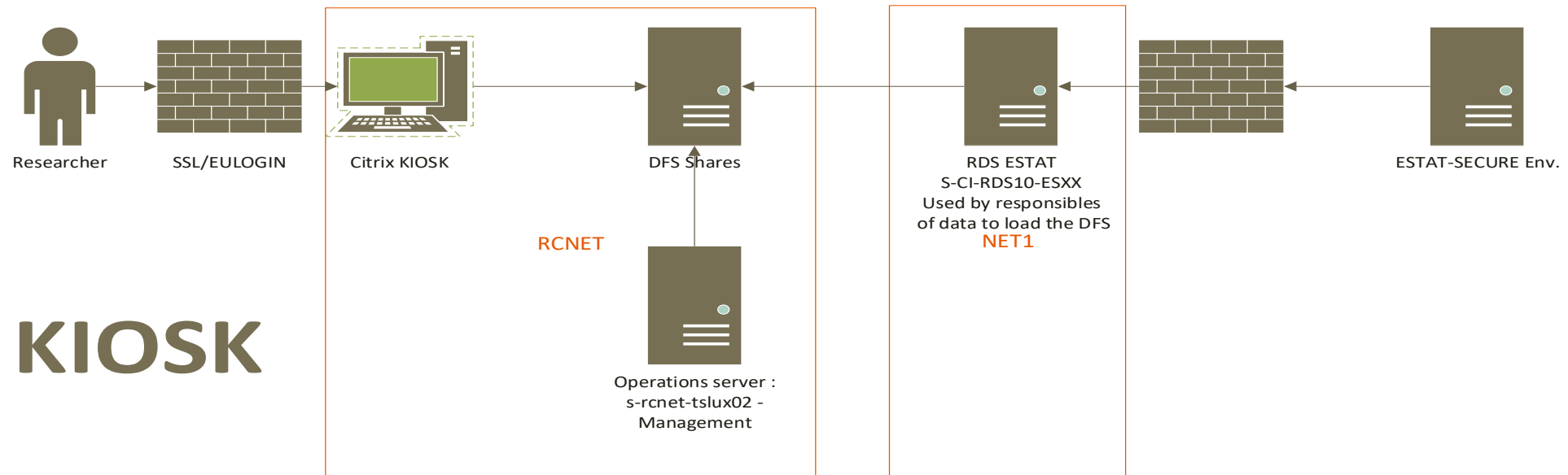
# Remote access to European microdata

## Implementation:

- Access via **accredited access points** at the premises of **recognised research entities**
  - Separate room for the operations (such as Eurostat's safe centre)
  - Access point compliant with organisational, technical and computer settings requirements
- Confidential data remains in Eurostat secure environment
- Access through DIGIT/Eurostat remote access system ('KIOSK')
- Output checking through Eurostat

# Remote access to European microdata

## Kiosk - architecture overview



# Remote access to European microdata

Current situation:

- For recognised research entities in **EU / GDPR area** and
- Secure use files of two domains (**CIS, SES** – same as at Eurostat safe centre)
- Several commercial and open source tools available in KIOSK:
  - STATA®, SAS®, R Studio, Tau Argus, Mu Argus, Office automation suite
  - ACRO tool (Automatic Checking of Research Outputs): proof-of-concept software based on STATA®
- Pilot implementation: JRC-ISPRA, first universities in the pipeline

# Remote access to European microdata

Possible future extension:

Inclusion of additional domains would require

- Agreement on (content of) secure use files for e.g. EU-LFS, EU-SILC,...  
-> should offer 'significantly' more than existing scientific use files
- Agreement of countries to allow access to the respective secure use files with their country data via remote access (in general / for certain projects)  
-> agreement of most if not all countries needed
- Eurostat resources for output checking
- ...

# Thank you



© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

