

Lisbon, 25-26 May 2023

LFS Methodology 2023

IT-LFS Microdata file under new EU regulation 1700/2019

Summary

- Microdata production in IT-LFS
- Types of microdata
- What's new with Eu 1700/2019
- Micro data for 2018-2020

IT Legal Basis for Personal Data Protection

Microdata file are released according to both current Italian regulations and the Institute's guidelines:

Legislative decree 6 September 1989, n. 322 Laws about National Statistical System an ISTAT, art. 6-bis (personal data processing), art. 9 (statistical confidentiality), art. 13 (statistical secrecy)

Legislative decree 30 June 2003, n. 196 Personal Data Protection Code

Legislative Decree 14 March 2013, n. 33 (Access for scientific purposes to data collected for statistical purposes)

DPR of 9 March 2022, Approval of National Statistical Programme 2020-2022

Different Users- Different Files

The results of the survey are also disseminated through the release of elementary data of different types, each oriented to a specific group of users:

File **SISTAN**

- Users: Entities of National Statistical System (Public Research Institutions, Ministries, Regions, Municipalities)
- NO anonymization, they contain all the validated information
- Statistical confidentiality issues are under the responsibility of the organization that access the data
- This file is provided also to ADELE (ISTAT Research Data Center) (disclosure risk is assessed by ISTAT researchers)

Different Users- Different Files

The results of the survey are also disseminated through the release of elementary data of different types, each oriented to a specific group of users:

File **SUF-MFR**

- Users: research entities by recognized Eurostat or Comstat
- Anonymization is foreseen:
 - Some variables are suppressed
NUTS5 codes, Household and individual direct identifiers
 - Other variables are recoded
Household typology, marital status, age, profession
- Provides a high level of information detail
- In the past it was called Standard file

Different Users- Different Files

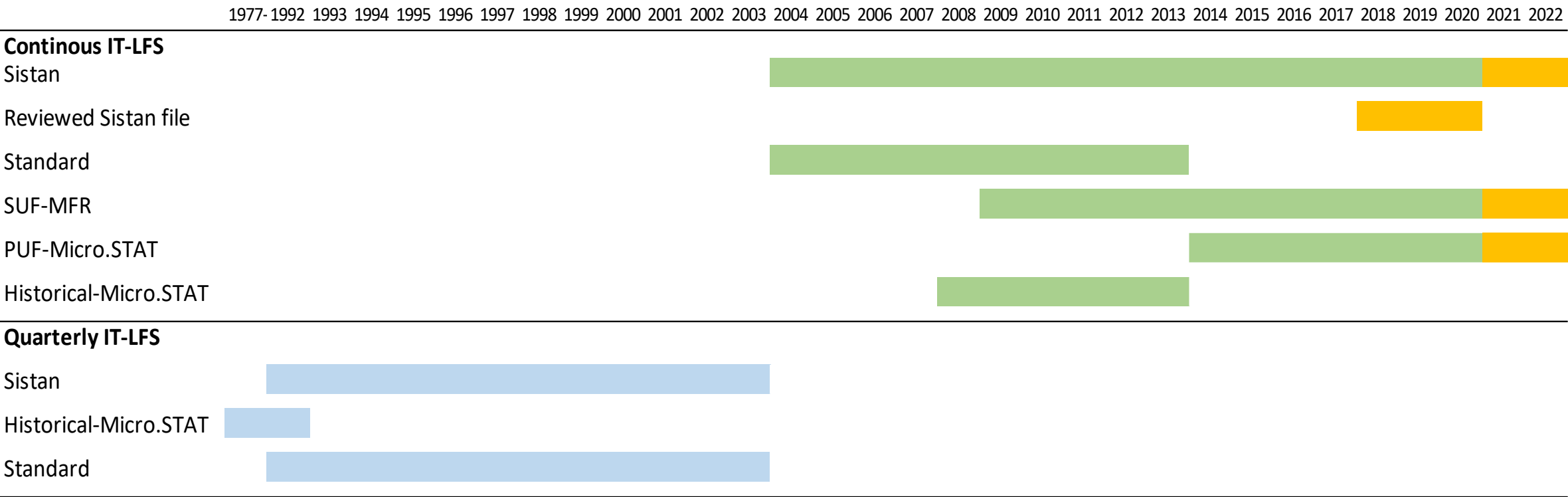
The results of the survey are also disseminated through the release of elementary data of different types, each oriented to a specific group of users:

File **PUF-mIcro.STAT**

- Users: freely available on ISTAT website
- Anonymization rules are stricter:
- Derived from SUF files, Historical editions
- For the period 2014-2020 they were defined by resampling
- From 2021 they are defined only through local suppression and global recoding
 - More consistent with published data
- Provides a lower level of information detail

Microdata file availability

Several structural breaks and innovations in the production of the LFS microdata have led to a rather complicated situation



Requests for accessing microdata files

March 2022- March2023

Type of Microdata File		
	Frequency	Percentage
Sistan	53	42,7%
SUF- Standard	17	13,7%
SUF-MFR	46	37,1%
ADELE	7	5,6%
Research Protocol	1	0,8%
TOTAL	124	

Guidelines for microdata production

A specific methodology is adopted for the preparation of the IT-LFS files for dissemination, in order to limit the risk of identification of the respondents.

- Initially all direct identifiers are removed
- We move on to the identification of the key variables, i.e. those which have at least one of the following characteristics:
 - rarity of the values in the population under study
 - visibility of the character, or some of its modalities, by an observer
 - traceability in external archives

For IT-LFS the main key variables to keep under control are: NUTS code of residence, work or study, the state of birth or citizenship, marital status or profession,

Guidelines for microdata production

To evaluate the identification risk, the sample frequencies of the combinations of each key variable by age group are first analyzed, and then the combinations of two or more key variables and at household level

$$\frac{n. individuals in distinct groups belonging to cells with frequency < k}{n. individuals in the data set} < p$$

$k_{MFR} \in \{2,3\}, p \in [0,0.1]$ for the SUF file

$k_{Micro} > k_{MFR}, p \in [0,0.01]$ for the public user file

When the previous condition is not satisfied, it is necessary to intervene using a specific statistical technique to protect the privacy of the respondents.

Detail of information

The statistical techniques adopted to protect the privacy of the respondents reduce the information content of the various microdata files:

	NUTS	Profession ISCO08	Economic Sector	Household Type	Age
SISTAN	NUTS5	5 digit	4 digit ATECO	41 levels	Single year
SUF-MFR	NUTS3	3 Digit	2 Digit ATECO	37 levels	0-2,3-5,6-10,11-14,15...74,75+
PUF-micro.STAT	NUTS2	1 Digit	12 Classes	7 levels	17 age groups

Reviewed microdata for 2018-2020

For the period 2018-2020, additional questions were included in the old questionnaires (the reason and duration of absence from work) to be able to classify the interviewees with the new definition of employed introduced by regulation 1700/2019. That allowed a link between the old definition of employment and the new one.

These reviewed data include main indicators, such as employment, unemployment, and inactivity rates, for socio-demographic characteristics such as gender, geographic areas, and age groups.

This backward revision allows us to produce microdata files with a reduced number of variables consistent with those released from 2021 onward.

To avoid anonymization issues they have been released as **SISTAN** microdata.

Reviewed microdata for 2018-2020

Information reconstruction was done for those who resulted employed according to the new definition but were not employed under the old one (about 0.05% of the total individuals).

Since most of the cases are women who have been on parental leave for more than three months, some information was imputed in a deterministic way, retrieving it from the section on “Previous work experience” or with Hot Deck methods.

In the opposite situation (about 0.09% of the total individuals), employed according to the old definition and unemployed according to the new, some information was stated as missing.

Tabella di cond3 per cond3_v				
cond3	cond3_v			Totale
	1	2	3	
1	58657	35	445	587051
	47.88	0.00	0.04	47.92
	99.92	0.01	0.08	
	99.51	0.06	0.08	
2	155	62558	0	62713
	0.01	5.11	0.00	5.12
	0.25	99.75	0.00	
	0.03	99.94	0.00	
3	2727	0	572564	575291
	0.22	0.00	46.74	46.96
	0.47	0.00	99.53	
	0.46	0.00	99.92	
Totale	589453	62593	573009	1225055
	48.12	5.11	46.77	100.00

Conclusion

With the release of microdata according to 1700/2019 an attempt was made to improve the quality of the data

- On the one hand we tried to ensure continuity and coherence with the previous series
- A significant improvement is the better coherence of the public user file with the official estimates:
 - balance the trade off between accessibility and reliability/accuracy
 - improve the timeliness of the administrative procedure for accessing SUF data

Grazie Thank you!

Alessandro Martini | alessandro.martini@istat.it