



EUROPEAN COMMISSION
EUROSTAT

Directorate B: Methodology, corporate statistical and IT services
Unit B-1: Methodology and corporate architecture

ESTAT/B1/WGSC(15)1
Available in EN only

7th meeting of the
Working Group on Statistical Confidentiality

Luxembourg, 30 November - 1 December 2015

Eurostat building BECH, Room Quetelet

Start: 30 November 2015, 13:00

End: 1 December 2015, 16:00

Item 3.2

Strategy for ESS public use files

STRATEGY FOR ESS PUBLIC USE FILES

1. OBJECTIVE

The objective of the document is to propose the strategy for the production of public use files in the European Statistical System.

The Working Group on Statistical Confidentiality is invited to discuss and endorse the proposed strategy.

2. BACKGROUND INFORMATION

Public use files (PUF) are defined in the Regulation (EC) No 223/2009 in the following way (Article 19):

"files consisting of anonymised records which have been prepared in such a way that the statistical unit cannot be identified, either directly or indirectly, when account is taken of all relevant means that might reasonably be used by a third party. If the data have been transmitted to the Commission (Eurostat) the explicit approval of the NSI or other national authority which provided the data is required."

Public use files are produced in many countries and are important source of data available to general public. Users use the PUF for the following reasons:

- To conduct statistical and scientific analysis
- To teach statistics / software use, to teach how to use record level data
- To get an idea of the structure and content of the more detailed confidential micro datasets available for the authorised users only.

In the past many requests for PUF have been addressed to Eurostat by researchers, academia and users of European statistics. However, production of PUF was not considered high priority for Eurostat because of the costs related with the development of PUF and unknown benefits.

Recently the experiences of some countries prove that production of public use files can be cost-effective if the standard methodologies and tools are used. At the same time the public use files can constitute a valid alternative for the users of scientific use files thereby contributing to the better cost-benefit ratio of the PUF production. Once produced, public use files are made available freely on the internet and there are no additional costs related with granting access.

A growing interest in the production of public use files in the European Statistical System (ESS) is reflected in the recent consultation of the WG on Statistical Confidentiality and Expert Group on Statistical Disclosure Control (EGSDC). Both groups were asked to indicate priority areas in the field of SDC where joint work co-financed by Eurostat would be considered beneficial. A harmonized methodology for public use files and scientific use files was ranked second in the resulting priority order.

Development of public use files is an important element of the strategy for access to microdata presented to the ESS Committee in September 2015. Notwithstanding some important comments on limited value of PUF due to anonymisation process, a lot of countries supported the idea of a harmonised approach to producing the files.

In December 2014 the specific grant agreement on the development of ESS public use files was signed with seven national statistical institutes (Netherlands-project coordinator, Austria, Finland, France, Germany, Hungary, and Slovenia). The objective of this project was to produce the PUF for the two most popular scientific use files: EU Labour Force Survey (EU-LFS) and EU Statistics on Income and Living Conditions (EU-SILC) for the participating countries. The details of this project can be found in the document under item 3.1 of the agenda.

The methodology for the development of PUF for the EU-LFS and EU-SILC was discussed at the meeting of the Expert Group on SDC in September 2015. One of the conclusions of this discussion was the need for a general strategy for the production of public use files for the ESS.

3. STRATEGY FOR ESS PUBLIC USE FILES

The strategy proposed in this document contains several elements that should be agreed upon before the public use files are made available for the users.

3.1. PUF coverage

The following table presents the datasets ordered according to the number of users of scientific use files since July 2013.

It is proposed to:

1. focus the production of PUF on the datasets that are the most popular amongst researchers. As far as the reference years are concerned, the PUF shall be produced for the most recent available years and maximum 10 past years.

Table 1 Use of scientific use files (8/07/2013 – 28/10/2015)

	Datasets	Number of users	Years available	Countries available
1	EU Statistics on Income and Living Conditions (EU-SILC)	381	2004-2013	EU, EFTA, RS
2	EU Labour Force Survey (EU-LFS)	300	1983-2013	EU, EFTA,
3	European Community Household Panel (ECHP)	82	1994-2001	EU-15
4	Community Innovation Survey (CIS)	72	2004, /6, /8, /10, /12	14 EU countries for 2012 data
5	Structure of Earnings Survey (SES)	60	1995, 2002, 2006, 2010	23 EU countries and NO (in 2010)
6	Adult Education Survey (AES)	38	2007, 2011	EU-28 (except HR), NO, CH, RS

7	European Health Interview Survey (EHIS)	27	2006-09	16 EU countries
8	Continuing Vocational Training Survey (CVTS)	14	2005, 2010	21 EU countries and NO
9	Community Statistics on Information Society (CSIS)	7	2008-2013	EU-28 (except DE), IS, NO
10	European Road Freight Transport Survey (ERFT)	5	2011	EU-28 (except MT), NO, CH

2. The public use files would be available for those countries that agree to the publication of their countries PUF for specific datasets (see item 3.3. below).
3. A common strategy for the production of national PUF for datasets for which Eurostat does not receive microdata (which are hence not available as ESS scientific use files) can also be investigated. Some members of the EGSDC suggested census data as a possible option. This will need to be further discussed.

3.2. Methodology for public use files

A methodology for public use files will be developed for each dataset separately. The methodology developed for EU-LFS and EU-SILC shall be a starting point for the other datasets, especially personal and household datasets. Ideally, the methodology shall be developed by the collaborative projects such as specific grant agreements.

3.3. Approval of the NSI on the publication of PUF

The PUF methodology developed by the collaborative (or other) projects shall be discussed and validated by the Expert Group on SDC and Working Group on Statistical Confidentiality (WG on Methodology). Each country may individually request some additional adaptations of the method and decide if the national data will be included in the ESS PUF. The PUF published by Eurostat will contain data of only those countries that agreed to be part of the ESS PUF. The decision of the countries will be requested for each dataset separately.

3.4. Publication of PUF

The data of the countries that agreed to be part of particular PUFs will be published on Eurostat website, most probably in the dedicated microdata access section: <http://ec.europa.eu/eurostat/web/microdata/overview>

Clearly, the countries may publish their own national PUF at the national level in parallel.

3.5. Conditions of use of PUF

The Expert Group for International Collaboration on Microdata Access of the OECD Committee on Statistics and Statistical Policy (CSSP) produced a report recommending ways of international access to microdata¹. Development of

¹ See: <http://www.oecd.org/std/microdata-access-final-report-OECD-2014.pdf>

public use files is strongly encouraged in the report as contributing to, amongst others, the increase of democracy (public access) and statistical literacy. The report contains an inventory of various practices regarding production and publication of PUF. In chapter 10 of the report entitled "Licensing Public Use Files" there are some considerations regarding publication of PUF, in particular the copyright and warranty issues are underlined. According to the report PUF may be published without any terms requiring an agreement or under licence. In the latter case it is recommended that before allowing the download of the dataset, the terms of the license are presented to the user.

The Eurostat's standard disclaimer is annexed to this document and can also be looked at here: http://ec.europa.eu/geninfo/legal_notices_en.htm. It is proposed to use the standard disclaimer for publication of ESS PUF.

4. NEXT STEPS

It is proposed to continue developing PUF according to the proposed calendar in the Table below.

Table 2 Proposed calendar for the development of the ESS PUF

What?	By?	How?	By when?
EU-SILC and EU-LFS			
Approval of methodology for EU-SILC and EU-LFS	WGSC	At the meeting on 31/11-1/12/2015	At the meeting or soon after the meeting
Consultation of the seven NSIs concerned (SGA(1) members) on the publication of EU-SILC and EU-LFS PUF on Eurostat website	NSIs participating in the SGA: AT, DE, FI, FR, HU, NL, SI,	In written	End of January 2016
Publication of the EU-SILC and EU-LFS PUF on Eurostat website (if approved by NSIs participating in the SGA(1))	Eurostat		End of February 2016
Application of the methodology developed for EU-SILC and EU-LFS PUF on the remaining countries and reference years	Eurostat	Specific contract under methodological framework contract	January 2017
Consultation of NSIs and publication of PUF	NSIs, Eurostat		First quarter 2017
Other datasets: ECHP, AES, CIS, SES			
Launch of the project on the development of	NSIs (interested FPA members)	SGA(3) to be signed	By the end of 2015

PUF for ECHP, AES, CIS, SES			
Development of the methodology for ECHP, AES, CIS, SES PUF	SGA(3) members	SGA(3)	May 2017
Validation of the methodology for ECHP, AES, CIS, SES PUF	EGSDC / WG on methodology	Meeting or in written	June 2017
Consultation of NSI participating in the SGA(3) and publication of the approved PUF	Eurostat/ NSIs participating in the SGA(3)		2017
Application of the methodology on other countries' data	Eurostat		2017
Consultation of remaining NSIs and publication of approved PUF	NSIs, Eurostat		2017

Production of PUF for other datasets is not foreseen at the moment (little demand for scientific use files).

The development of PUF for Census data will be further discussed with the Eurostat responsible unit and Census WG.

5. CONCLUSIONS

The WGSC is invited to discuss the proposed strategy for PUF and to provide comments on the draft calendar for production of ESS PUF.

The WGSC (WG on methodology in the future) will be kept informed of all relevant developments in this area.

Annex 1 Standard Eurostat's disclaimer for the information published on Eurostat website.

Legal notice

The information on this site is subject to a [disclaimer](#), a [copyright notice](#) and rules related to [personal data protection](#).

Disclaimer

The European Commission maintains this website to enhance public access to information about its initiatives and European Union policies in general. Our goal is to keep this information timely and accurate. If errors are brought to our attention, we will try to correct them.

However the Commission accepts no responsibility or liability whatsoever with regard to the information on this site.

This information is:

- of a general nature only and is not intended to address the specific circumstances of any particular individual or entity;
- not necessarily comprehensive, complete, accurate or up to date;
- sometimes linked to external sites over which the Commission services have no control and for which the Commission assumes no responsibility;
- not professional or legal advice (if you need specific advice, you should always consult a suitably qualified professional).


Please note that it cannot be guaranteed that a document available on-line exactly reproduces an officially adopted text. Only the Official Journal of the European Union (its printed edition, or since 1 July 2013 its electronic edition made available on the EUR-Lex website), is authentic and produces legal effects.

It is our goal to minimize disruption caused by technical errors. However some data or information on our site may have been created or structured in files or formats that are not error-free and we cannot guarantee that our service will not be interrupted or otherwise affected by such problems. The Commission accepts no responsibility with regard to such problems incurred as a result of using this site or any linked external sites.

This disclaimer is not intended to limit the liability of the Commission in contravention of any requirements laid down in applicable national law nor to exclude its liability for matters which may not be excluded under that law.


Copyright notice

© European Union, 1995-2015

Reuse is authorised, provided the source is acknowledged. The reuse policy of the European Commission is implemented by a [Decision of 12 December 2011](#)  [728KB] .

The general principle of reuse can be subject to conditions which may be specified in individual copyright notices. Therefore users are advised to refer to the copyright notices of the individual websites maintained under Europa and of the individual documents. Reuse is not applicable to documents subject to intellectual property rights of third parties.

Personal Data Protection

The European Union is committed to user privacy. The policy on "protection of individuals with regard to the processing of personal data by the Community institutions" is based on [Regulation \(EC\) N° 45/2001](#)  [234 KB] of the European Parliament and of the Council of 18 December 2000. This general policy covers the European Union's family of institutional Web Sites, within the eu domain. Although you can browse through most of these Web Sites without giving any information about yourself, in some cases, personal information is required in order to provide the [e-services](#) you request. Web Sites that require such information treat it according to the policy described in the Regulation mentioned above and provide [information about the use of your data](#) in their specific privacy policy statements.

In this respect:

- For each specific e-service, a controller determines the purposes and means of the processing of personal data and ensures conformity of the specific e-service with the privacy policy;
- Within each Institution, a Data Protection Officer ensures that the provisions of the Regulation are applied and advises controllers on fulfilling their obligations (see art. 24 of the Regulation);
- For all the Institutions, the [European Data Protection Supervisor](#) will act as an independent supervisory authority (see art. 41 to 45 of the Regulation).

The European Union's family of institutional Web Sites, within the eu domain, provides links to third party sites. Since we do not control them, we encourage you to review their privacy policies.

What is an e-service?

An e-service on EUROPA is a service or resource made available on the Internet in order to improve the communication between citizens and businesses on the one hand and the European Institutions on the other hand.

Three types of e-services are or will be offered by EUROPA:

1. **Information services** that provide citizens, media, business, administrations and other decision makers with easy and effective access to information, thus increasing transparency and understanding of the policies and activities of the EU;
2. **Interactive communication services** that allow better contacts with citizens, business, civil society and public actors thus facilitating policy consultations, and feedback mechanisms, in order to contribute to the shaping of policies, the activities and the services of the EU;
3. **Transaction services** that allow access to all basic forms of transactions with the EU, e.g. procurement, financial operations, recruitment, event enrolment, acquisition or purchase of documents etc.

Information contained in a specific privacy statement.

A specific privacy policy statement will contain the following information about the use of your data:

- What information is collected, for what purpose and through which technical means the EU collects personal information exclusively to the extent necessary to fulfil a specific purpose. The information will not be re-used for an incompatible purpose;
- To whom your information is disclosed. The EU will only disclose information to third parties if that is necessary for the fulfilment of the purpose(s) identified above and to the mentioned (categories of) recipients. The EU will not divulge your personal data for direct marketing purposes;
- How you can access your information, verify its accuracy and, if necessary, correct it. As a data subject you also have the right to object to the processing of your personal data on legitimate compelling grounds except when it is collected in order to comply with a legal obligation, or is necessary for the performance of a contract to which you are a party, or is to be used for a purpose for which you have given your unambiguous consent;
- How long your data is kept. The EU only keeps the data for the time necessary to fulfil the purpose of collection or further processing;

- What are the security measures taken to safeguard your information against possible misuse or unauthorised access;
- Whom to contact if you have queries or complaints.

How do we treat e-mails you send us?

Many web pages on Europa have a "Contact" button, which activates your e-mail software and invites you to send your comments to a specific functional mailbox.

When you send such a message, your personal data is collected only to the extent necessary to reply. If the management team of the mailbox is unable to answer your question, it will forward your e-mail to another service. You will be informed, via e-mail, about which service your question has been forwarded to.

If you have any questions about the processing of your e-mail and related personal data, do not hesitate to include them in your message.