

UNITED NATIONS ECONOMIC
COMMISSION FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN
STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Helsinki, Finland, 5 to 7 October 2015)

Topic (iv): Access to Statistical Data for Scientific Purposes

Access to EU microdata for research purposes

Aleksandra Bujnowska *

* Methodology and corporate architecture unit, Eurostat, Luxembourg, e-mail:
Aleksandra.Bujnowska@ec.europa.eu

Abstract: This paper reports on the implementation of the new legal framework for access to EU microdata for research purposes and presents the ongoing and planned work in this area. A similar report was presented to the European Statistical System Committee meeting in September 2015.

1 Introduction

Statistical confidentiality is a fundamental principle of official statistics enshrined in the Treaty¹ and in the European statistics Code of Practice. Harmonisation of principles and guidelines as regards protection of confidential data is the obligation of Eurostat and national statistical authorities in the European Statistical System.

Confidential data are not only the basis for production of statistics but also considered as indispensable source of information by researchers. Access to microdata for scientific purposes is growing in importance and allows researchers to address many questions in the fields of economic and social sciences.

Access to microdata at the EU level allows researchers to carry out comparative analysis on the basis of the harmonised European data.

2 Microdata at the EU level

Countries transmit data to Eurostat for the production of European statistics. In some domains data are sent at micro level and in other at aggregated level (tables).

¹ Article 338 of the Treaty on the functioning of the European Union.

Eurostat may grant access to microdata transmitted for the production of European statistics. This is subject to the agreement of the countries providing the data.

EU microdata are available in two forms:

- “scientific-use files”: anonymised data delivered to researchers on electronic devices (CD-Rom, DVD, etc.);
- “secure-use files” available in Eurostat's "safe centre” in Luxembourg (only de-identified data);

Scientific use files are especially prepared to make the identification of survey respondents more difficult. Statistical disclosure control (SDC) methods are applied to this data to reduce to an appropriate level and in accordance with best practices the risk of identification of the statistical unit.

Methods of protection are decided in collaboration with national statistical authorities, taking into account the mode of access, the probability of re-identification, utility, harmonisation and the impact of unlawful disclosure.

The actual application of the SDC methods is performed by Eurostat (in-house or subcontracted) or by the statistical authority providing the data.

Scientific use files are available for the following data collections:

- European Community Household Panel (ECHP);
- Labour Force Survey (LFS);
- European Union Statistics on Income and Living Conditions (EU-SILC);
- Adult Education Survey (AES);
- Community Innovation Survey (CIS);
- Structure of Earnings Survey (SES);
- European Health Interview Survey (EHIS);
- European Road Freight Transport Survey (ERFT);
- Community Statistics on Information Society (CSIS);
- Continuous Vocational Training Survey (CVTS);

SDC methods are not applied to secure use files. These files can be made available to researchers only in the secure environment ensuring that the results of the statistical analysis are not released prior to output checking.

Secure use files are available in the Eurostat safe centre (on-site) in Luxembourg for:

- Community Innovation Survey (CIS);

- Structure of Earnings Survey (SES);
- NEW: Linked micro-aggregated data on ICT usage, innovation and economic performance in enterprises (Micro-moments dataset).

3 Access to EU microdata

Eurostat grants access to microdata of the European Union (EU) Member States (MS) and some other European countries (EEA/EFTA countries and some EU-candidate countries) on the basis of the new Regulation that entered into force in July 2013².

The access procedure is divided in two steps:

Step 1 – Recognition of the organisation as a research entity;

Step 2 – Submission of the research proposal

3.1 Step 1 Recognition of the organisation as a research entity

The recognition of research entities is a new procedure introduced by the new Regulation. Entities are recognised only once and can subsequently submit as many research proposals as they wish. Nevertheless, this procedure takes time (around 4 weeks as very often additional information must be requested or certain points of applications clarified). The following criteria must be fulfilled by the research entities:

- The purpose (mission, statute) of the entity shall refer to research;
- The entity's list of publications and research projects must be presented;
- The entity must be independent in formulating scientific conclusions
- The entity must have adequate security safeguards.

Since July 2013 over 450 entities has been recognized by Eurostat as research entities (see statistics below). The application is done via Eurostat website³.

The assessment of applications shows that pure research entities, such as universities, are increasingly complemented by entities where research is carried out together with other activities. In such cases, the Regulation requires that the research department is identified and recognition is limited to that department and not to the entire organisation. According to the Regulation, one of the prerequisites for the recognition is that the research department must be independent from other areas of the organisation, a clause that is sometimes difficult to assess.

These types of bodies with mixed (research and other) objectives include:

- Commercial organisations (consultancy firms doing also research, contractors of the European Commission);
- International organisations;

² Commission Regulation t (EU) No 557/2013 on access to confidential data for scientific purposes.

³ Eurostat website (<http://ec.europa.eu/eurostat/>) Data / Access to microdata. The full list of recognized research entities is also published there.

- Research departments of non-research bodies like central banks, ministries, Commission DGs.

Recognized research entities by countries

Member States		EFTA countries	
Austria	19	Island	1
Belgium	19	Norway	9
Bulgaria	3	Switzerland	11
Croatia	3		21
Cyprus	1		
Czech Republic	7	EU-candidate countries	
Denmark	5	Macedonia	1
Estonia	5	Serbia	1
Finland	5	Turkey	6
France	28		8
Germany	53		
Greece	9	Other countries	
Hungary	9	Australia	4
Ireland	8	Canada	6
Italy	38	Israel	3
Latvia	1	South Korea	5
Lithuania	2	Tunisia	1
Luxembourg	4	Ukraine	1
Malta	1	USA	23
Netherlands	20		43
Poland	10		
Portugal	13	EU institutions, bodies and agencies	13
Romania	3		
Slovakia	6		
Slovenia	2	International organisations	3
Spain	40		
Sweden	5		
United Kingdom	50		
	369	TOTAL	457

3.2 Step 2 Submission of the research proposal

In the research proposal submitted to Eurostat researchers working at the recognised research entities describe:

- research project;
- statistical methods to be applied;
- the data and variables to be used;

- why access to microdata is necessary for the project;
- how the results of research will be published;
- how the confidential data will be securely stored at the premises of the research entity.

National statistical authorities (NSA) that provided the data requested are consulted on each research proposal. If the NSA refuses the access, the data of the country is removed from the microdata file.

Under new legal framework the Eurostat microdata access team received 640 research proposals. The overview of the requests by datasets is presented in table 2. Almost 70% of access requests account for EU Statistics on Income and Living Conditions (EU-SILC, 40%) and EU Labour Force Survey (EU-LFS, 30%). The most popular domains of research include: labour market analysis, social studies, innovation, growth and social impact of crisis.

Table 2 Number of research proposals received by datasets (in one application several datasets can be requested)

	ECH P	EU- LFS	EU- SILC	AES	CIS	SES	EHIS	ERFT	CVTS	CSIS	MMD	
2013	7	38	45	6	13	7	3	1	0	0		120
2014	41	135	164	21	31	32	13	1	9	2		449
2015 *	22	95	127	8	22	12	6	2	3	5	2	304
	70	268	336	35	66	51	22	4	12	7	2	873

*Requests received until end of July 2015.

4 Next steps

In the short term, Eurostat plans to improve current modes of access, especially data transmission (e.g., dismiss the transmission of CDs/DVDs in favour of more modern solutions such as FTP store), and to launch a system for management of applications for microdata access (workflow system). This should reduce the time of processing of the microdata access requests. It is also essential to reinforce support services related to microdata access (trainings, metadata services, user support). These services might be provided in collaboration with data archives which specialise in such activities.

4.1 Workflow system

Eurostat is developing a system that will facilitate (and digitalise) management of microdata access requests. The system is based on webforms to be used by researchers to introduce the information about their projects. National statistical authorities will be consulted on the new research proposals via this system as well. The system should be operational by end 2015.

4.2 Public use files (PUF)

In the context of the activities of the Centre of Excellence on Statistical Disclosure Control, Eurostat has launched a specific grant agreement aiming at development of PUFs mostly for testing and training purposes. The project concerns EU-SILC and LFS data. The first PUFs shall be available at the beginning of 2016. PUFs can represent a viable alternative to the request of scientific use files in some cases⁴. Development of PUFs for other datasets is also planned.

4.3 Self-study material on the use of confidential data

Eurostat is developing a self-study package that will be recommended to all users of EU microdata. It will explain why confidential data needs to be protected and how to ensure security of the files. The material will also include a test allowing researchers to assess their knowledge in this specific domain.

4.4 Collaboration with data archives

Services and support for the research community can be improved and made more efficient through collaboration between the European Statistical System and Consortium of European Social Science Data Archives (CESSDA)⁵. The Data without Boundaries project (DwB)⁶ has identified potential areas of such collaboration. The partnership could be established once CESSDA becomes European Research Infrastructure Consortium (ERIC). It could aim initially at providing some user support services, and potentially be extended to some data supply services. The extent of the collaboration depends upon CESSDA's capacity and priorities and also on the domains the ESS wishes to deliver through such potential partnership.

The possible partnership could be focused on the following services⁷:

- Metadata products and services

MISSY⁸ is an online service platform that provides systematically structured metadata for official statistics. The system was developed by CESSDA member GESIS. MISSY consists of a metadata editor and a web-based metadata content publishing tool. The service is offered over the Internet and is free of charge. Recently, with the support of

⁴ See more on the PUF project in the separate conference paper submitted to session II.

⁵ <http://cessda.net/eng>

⁶ Data without Boundaries was funded by FP7 budget and its mission was to "support equal and easy access to the rich resources of official microdata for the European Research Area, within a structured framework where responsibilities and liability would be equally shared. During its four-year lifespan the DwB worked towards preparing a comprehensive European service with better and friendly metadata, a more harmonized transnational accreditation and a secure infrastructure that would allow transnational access to the highly detailed and confidential microdata, both national and European, so that the European Union would be able to continuously produce cutting-edge research and reliable policy evaluations" (<http://www.dwbproject.org/>). DWB project ended on 30 April 2015.

⁷ Part co-written with CESSDA.

⁸ <http://www.gesis.org/missy/eu/missy-home>.

DwB, MISSY has been expanded to document Eurostat micro-data⁹. MISSY is fully DDI compatible and is currently the only information system documenting Eurostat data in a systematic, machine-readable way down to the variable level.

To keep MISSY a living system it has to be updated regularly with the yearly updates of LFS and SILC as well as with the updates of the other surveys occurring less often.

- Researcher training and learning in European Statistics microdata

Through the Data Without Boundaries project a series of training courses has been developed for the LFS, SILC cross-sectional, SILC longitudinal, and AES sources. The trainings were hosted by several CESSDA partners and NSIs and offered for the first time hands-on training with real microdata from Eurostat. This was made possible by the permission of a number of NSIs to use subsamples of their data for these events.

Training courses can improve the information available to researchers about access procedures, content of the data, research options, and services provided around these data.

- Communication of research accreditation requirements, and user support through the process

CESSDA members can offer local support for research institutions and research projects seeking Eurostat accreditation. Depending on the nature of the collaboration, this could be limited to simply providing the accreditation rules on the local websites, through to offering assistance for form completion and providing authenticated research entity information.

- Resource discovery and data catalogue

As part of collaboration it is proposed that the ESS microdata products (scientific use and secure use files, and any public use files) should be included in the Data Catalogue and discoverable through the Resource Discovery Portal managed by CESSDA. Researchers would find ESS data where they are relevant to their searches; and the metadata, and the procedures for gaining accreditation and access, can be made available through the service.

- User conferences

CESSDA is considering the future of the European Data Access Forum, which is an initiative of Data Without Boundaries. The Forum brings together NSIs, data service providers, and data users every two years. The Forum is an excellent occasion to consider user needs and priorities, and to consult on changes and innovations in research data access. A collaborative approach could result in future European Data Access Forums being jointly hosted by CESSDA and Eurostat.

⁹ In January 2015 MISSY went live for EU-LFS, EU-SILC and AES. The SES and CIS followed in April 2015 and EHIS metadata in June.

- Impact and benefits realisation

CESSDA has policies and practices in place to capture the benefits of the research projects enabled by data access. The task is not easy, and depends on good practices in the research community, good governance of the conditions of data licenses, and other solutions such as proper citation and permanent identifiers. Collaboration with the ESS could deliver a common expectation on the research community for proper citation and reporting on publications, and also on the use of tools that make this task easier to perform.

5 Conclusions

With the support of the ESS Committee to the proposed ways of improvements of microdata access services, Eurostat will continue the deployment of the identified actions and could start exploring and piloting the practical details of a potential collaboration with CESSDA (once the ERIC status is confirmed).