

WP.43
ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)**

EUROPEAN COMMISSION

CONFERENCE OF EUROPEAN STATISTICIANS

**STATISTICAL OFFICE OF THE EUROPEAN
UNION (EUROSTAT)**

Joint UNECE/Eurostat work session on statistical data confidentiality
(Tarragona, Spain, 26-28 October 2011)

Topic (viii): International projects, groups and forum dealing with data access, release and related methodologies

The Future of Access to European Confidential Data for Scientific Purposes

Prepared by Aleksandra Bujnowska and Jean-Marc Museux, Eurostat

The Future of Access to European Confidential Data for Scientific Purposes

Aleksandra Bujnowska and Jean-Marc Museux

Methodology and Research Unit, Eurostat, European Commission, L-2920, Luxembourg, e-mail: Aleksandra.Bujnowska@ec.europa.eu, Jean-Marc.Museux@ec.europa.eu

Abstract: European statistics are usually based on national data collected and produced by the national statistical authorities of the Member States. Regulation (EC) 223/2009 on European statistics establishes a legal framework for the development, production and dissemination of European statistics. Article 23 of that Regulation provides a framework for access to European confidential data - which only allow indirect identification of the statistical units - by researchers carrying out statistical analyses for scientific purposes. The Regulation also stipulates that the approval of the national authority which provided the data is required if the data have been transmitted to Eurostat. Eurostat's objective is to establish a future system of access to confidential data for scientific purposes which is consistent with the vision of the method of producing European statistics. In this system the confidential data would be available at various access points in the Member States, and the national authorities will be empowered to grant access to the entire set of EU data.

1 Introduction

Access to EU confidential data for scientific purposes began in 2002 with the entry into force of Commission Regulation 831/2002, which established the procedures for access, i.e. the conditions under which access to confidential data transmitted to the Commission (Eurostat) may be granted and the rules of cooperation between the Commission and national authorities in order to facilitate such access.

The Regulation defines the research bodies that may request access, the component parts of the research proposal, accessible datasets and modes of access. The procedure consists of two stages. In the first stage, Eurostat assesses whether the body requesting access can be considered as a research body, and in the second stage the research proposal is analysed. The national authorities that transmitted the data are consulted on each research proposal.

The researcher may have access to anonymised microdata available on CD-ROM or to the confidential data without direct identifiers in the Eurostat safe centre (in Luxembourg).

Figure 1 presents the process of transferring the data from national authorities via a secure transmission channel (eDAMIS) and the two modes of access to confidential

data for scientific purposes, namely: on-site access in Luxembourg and off-site access to anonymised microdata available on CD-ROM.

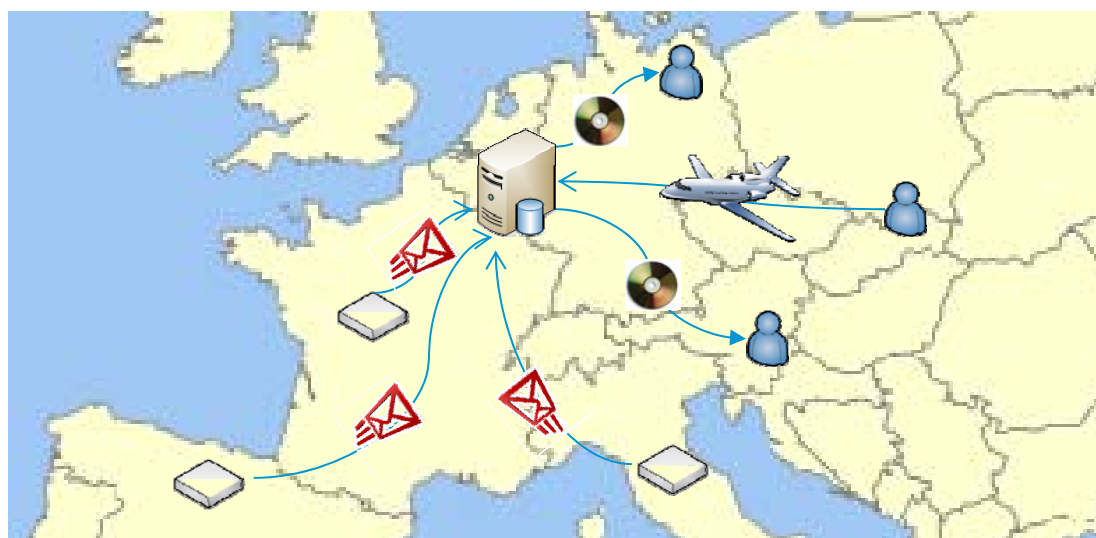


Fig 1. The current business process of collecting the microdata from Member States and providing access to confidential files compiled by Eurostat for scientific purposes.

2 Eurostat's vision of the future of access to European confidential data for scientific purposes

European statistics are developed, produced and disseminated on the basis of uniform standards and harmonized methods. The National Statistical Institutes (NSIs) of the Member States collect and produce harmonized data which are compiled by Eurostat in order to construct statistics at EU level.

In accordance with Eurostat's "vision", all the data collected under European legislation should be considered as the common property of the ESS and - consequently - any national authority should be empowered to grant access to all the European data, as long as certain commonly agreed basic principles are met. The targeted policy would be a "Schengen-like" microdata access policy, whereby it would be possible to access the whole sets of European data from any accredited entry point, with the authorization, management and control of access being delegated at the appropriate national/international level. Basic principles and standards should cover the full range of access modes: from access to non

anonymised data in the physical premises of research data centers to the release of anonymised datasets, including via different kinds of remote access depending on the local infrastructure. The Schengen-like microdata access in the EU would consist of:

- a unique accreditation mechanism for institutions and researchers accessing EU datasets,
- a distributed database with local versions of confidential data sets prepared by NSIs, with credentials being set locally,
- a central directory of files and access maintained by Eurostat,
- access to the network via a terminal server solution (remote access technology).

In the process of implementing the new vision, it is vital to ensure the appropriate level of compatibility with national provisions with respect to protection of statistical confidentiality. The security and integrity of the systems and of the data is an overarching principle at every stage of the process.

In the new **business process**, the NSIs will:

- 1) control and manage access at the national level; in particular, NSI staff will assess the research entity and research proposals on behalf of the ESS partners (safe people, safe project);
- 2) prepare national microdata according to the mode of release requested in the research proposal and in line with common requirements established at the EU level (safe data);
- 3) manage and supervise access to the confidential data in the local safe settings. Standard requirements for accountability, output checking, logical and physical security of systems (safe settings) will be laid down.

The comparison between the current and the targeted business architecture is set out in Figure 2.

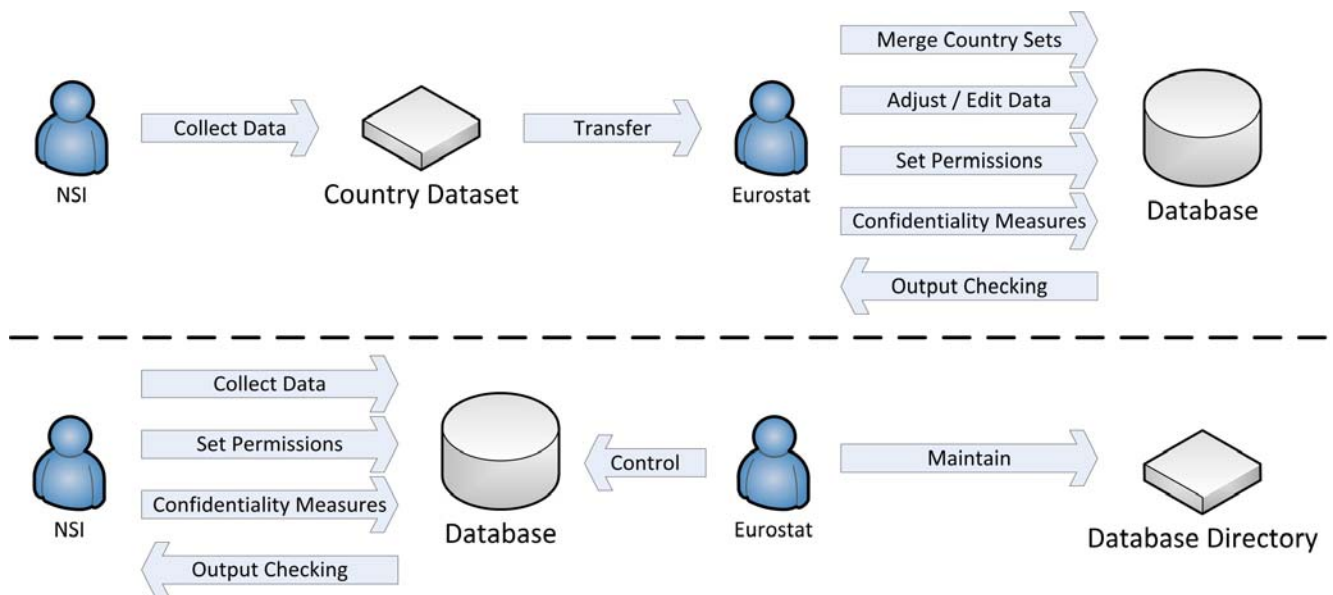


Fig 2 The comparison between the current and the targeted business architecture.

In order for it to be implemented, this European "vision" requires national authorities to accept changing the perception of exclusive ownership of the data to common ownership and shared responsibilities in respect of the EU data. In the long term, the existing legal framework will probably have to be changed in order to make these fundamental principles explicit. A careful cost benefit analysis must also be developed, as well as an agreed model for cost/burden sharing. A smooth step-by-step transition should take place.

3 Implementation

The "vision" presented above can be regarded as a long-term objective of Eurostat. In the short-to-medium term, Eurostat aims to widen the access to EU microdata as set out in Regulation No 223/2009 on European statistics. In particular, it is planned to decentralise access. Projects on the design of future access infrastructure have been launched. Changes in the legal framework are also being implemented, as well as sophisticated methodological solutions, which allow the protection of microdata in the distributed environment.

It is expected that implementation of the recently launched projects (details below) will result in the creation of a system of national safe centres, where access will be provided to the confidential data of the participating countries. Initially, only a few countries will take part, but more countries are expected to join in the coming years. The researchers will therefore not need to physically go to Eurostat in order to have

access to confidential data, because they will be able to use the local access facilities. The NSIs will check the results of the scientific analysis carried out in their safe centre, regardless of the source of the data.

Researchers will also have access to anonymised microdata via a network of data providers which will not only provide access to microdata, but will also support the user, and offer the necessary guidance on how to make use of the data.

NSIs will be directly involved in the production of anonymised microdata sets. They will also check the results produced by researchers in the safe centre on behalf of the countries whose data are used.

Even though the access will be distributed, there will be one reference database where all confidential data will be stored. It will be accessible through remote secure channels.

The necessary access procedures and access protocols will be established and will be implemented by the ESS partners. Eurostat's role will be one of coordination and control.

3.1 Access infrastructure

Two ESSnet projects have been launched with the aim of designing the access infrastructure. The first such project - a feasibility study on decentralized access to EU microdata - started in 2009 and brought together five countries, namely: DE (project leader), IT, UK, NL and HU. The second ESSnet starts in 2011 and will implement and test the solutions proposed by the first project team. This pilot on limited network infrastructure provides for access to the data of the countries participating in the project. The researchers will access the data remotely using a terminal server solution from the NSI data centres (UK, FR, HU, DE, PT). The data will be stored in the specific secure environment managed by Eurostat (SICON).

The project team will:

- study the feasibility and cost benefit of the extension of the network,
- refine the cost model
- fine-tune procedures and standards for decentralization
- work on output checking
- supervise and support researchers using the infrastructure.

The two ESSnets projects focus on devising solutions for the European Statistical System. The 7th Framework Programme project - Data without Boundaries (DWB) - works on the broader infrastructure, which includes not only NSIs but also data archives and research bodies.

The purpose of the DWB project is to enhance researchers' access to national and European official microdata. Its aim is to bring together NSIs, data archives and researchers to agree on standards and common views, and to build mutual trust. The project will enable researchers to test the access infrastructure and to gain access to other countries' data.

3.2 Revision of Regulation 831/2002 on access to confidential data for scientific purposes

The current legal framework on access to EU confidential data for scientific purposes is being revised. It needs to be adapted to the new Regulation on European statistics and to take account of the changes in the administrative, technical and methodological environments.

The weak points of the current method of access are the following:

1. only a limited types of bodies are allowed access; in practice access is restricted to European universities and research bodies; other research bodies must undergo a lengthy accreditation process;
2. only anonymised data can be released and transferred to research institutions, and access to original data is provided only in Eurostat's safe centre;
3. the datasets available are limited in scope – at present there are six surveys available as anonymised microdata files and two are ready to be used in the Eurostat safe centre.

Eurostat is now working on the revised version of the Regulation in collaboration with NSIs. Its main principles are:

- Accreditation not limited to EU universities and research bodies
- New modes of access (remote access) enabled
- Involvement of external partners (e.g. data archive) in the provision of access to researchers
- New and cost effective procedures
- Flexibility in incorporating new standards.

The revised Regulation is currently being discussed in the ESS, and there are plans for its possible entry into force in 2012.

3.3 New methodological solutions

The project team of the ESSnet on common tools and harmonized methodology for the SDC in the ESS is working on the last part of the structure – methodological solutions. As far as microdata access is concerned, the ESSnet will develop a coordinated approach to protection of the EU data files. It will allow the NSIs to

apply their own rules, while ensuring consistency and comparability between the data from the Member States.

The future projects will focus on standardization of methods in the ESS. Guidelines will be developed to recommend solutions for microdata protection at the ESS level.

4 Conclusions

Provision of access to EU microdata sets is one of Eurostat's core activities. At present, the system is managed centrally in Luxembourg. The trend towards a more decentralized model, in which data providers at the national level are directly involved in granting access to EU confidential data, is ongoing. Changes in legal, technological and methodological frameworks are being made in order to develop a modern system for the provision of access to microdata in the European Statistical System.