

Working Paper
ENGLISH ONLY

**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
(Ottawa, Canada, 28-30 October 2013)

Topic (iii): Modes of access to microdata

Improvement of access to European microdata

Prepared by Maurice Brandt, Federal Statistical Office, Germany

Improvement of access to European microdata

Maurice Brandt*

* Federal Statistical Office, Research Data Centre, Gustav-Stresemann-Ring 11, 65189 Wiesbaden, Germany, e-mail: maurice.brandt@destatis.de

Abstract: To improve the access to European microdata for scientific purposes the ESSnet project “Decentralised and Remote Access to Confidential Data in the ESS” (DARA) was conducted. The aim of the ESSnet-project DARA is to establish a secure channel from a safe centre within a National Statistical Institute (NSI) to the safe server at Eurostat so that researchers can use confidential EU microdata in their own Member States without travelling to Luxembourg.

The ESSnet DARA-project team has defined a concept of technical implementation and safety requirements for a remote access system. The concrete task of participating NSIs is to provide a secure channel to guarantee access for data users to the central node and also to provide service and IT-support for the researchers on the local national level.

The project team has drafted a handbook with descriptions and guidelines for NSI staff and researchers and an accreditation system for access facilities. For a prove of the concept and feasibility, the project team has implemented a remote access pilot with 6 access points in 5 countries.

1 Introduction

A major task of the European Statistical System (ESS) is to provide EU institutions, Member States and the public with reliable information about the society, economy, environment and development in the European Union. Therefore it is necessary to collect, process and analyse statistical data of the Member States. It is an elaborate procedure to harmonize the questionnaires, the preparation process and the data delivery of 28 Member States in the EU. It is necessary to achieve comparable information between Member States and useful aggregated values for the entire EU. In consequence of this, these data represent a precious and unique data source for the international research community. The full potential of these data is not exhausted with the publication of standardized aggregated tables. Furthermore there is a lot more analysis potential to model the European integration process, differences between Member States or best practice examples with multivariate and sophisticated evaluation.

The access to these confidential data can be provided for scientific purposes by sets of anonymised microdata or in the premises of Eurostat in the Safe Centre in Luxembourg. The anonymised microdata are a valuable source and their advantage is, that they can be used in the researchers own institutions. But for a lot of researchers the data sets are not detailed enough to conduct their analysis. They

prefer to access the original microdata (without direct identifiers). By now there is only the one possibility for European researchers to travel to Eurostats Safe Centre in Luxembourg what means a heavy burden, caused by the local constraints.

The ESSnet project **“Decentralised and Remote Access to Confidential Data in the ESS” (DARA)** deals with the implementation of a remote access system from Safe Centres in the National Statistical Institutes (NSI) of the Member States (MS) to the EU statistics at Eurostat. The goal of the ESSnet DARA project is to provide the basis for a European microdata access system for scientific purposes. With such a kind of system researchers do not have to travel to Eurostat in Luxembourg anymore; they can access detailed confidential microdata from their own Member States via remote access connection whereas the microdata itself remain at a secure place.

An easier access to use European microdata could lead to new empirical findings and knowledge about the “European society” in economics, social sciences and environmental sciences. The experiences in several Member States show, that the infrastructure of microdata access will be heavily used, once it is established in a convenient way for researchers.

The ESSnet DARA project was accompanied by the revision of the regulation to access confidential data of the EU statistics for scientific purposes (Reg. (EC) No. 831/2002). The new regulation was adopted and is in charge since July 2013 under the Commission Regulation (EU) No. 557/2013. The new regulation shall ease the access to EU microdata from a legal point of view and shall lay the basis to a more decentralised access in Europe.

The project has started in October 2011 with a duration of 2 years and was conducted with the following partners:

Groupe des Écoles Nationales d'Économie et de Statistique, GENES

Hungarian Central Statistical Office, Hungary, HSCO

Office for National Statistics, United Kingdom, ONS

Instituto Nacional de Estatística, Portugal, INE

State Statistical Institute Berlin-Brandenburg, Germany, AfS Berlin

Federal Statistical Office Germany (co-ordinator), DESTATIS

2 Project Content

The ESSnet project “Decentralised and Remote Access to Confidential Data in the ESS” (DARA) was the follow up project of the previous ESSnet project “Decentralised Access to EU Microdata Sets” (DA) which had proven the feasibility to access confidential data through a remote desktop connection in the Safe Centres of the Member States. One outcome of this project was the decision for a remote access system and as a central result the recommendation for the implementation of a pilot study was given.

The aim of the ESSnet project DARA is to establish a secure channel from a safe centre within a NSI to a secure server at the European Commission’s network so that researchers can use confidential EU data in their own Member States.

The ESSnet DARA project team has defined a concept of technical implementation and safety requirements for a remote access system. The usability of this system, the capability to analyse data and the appropriate safety requirements will be implemented by Eurostat. The concrete task of participating NSIs is to provide a secure channel to guarantee access for data users to the system at Eurostat and also to provide service and IT-support for the researcher on the local national level.

One important part of the ESSnet DARA is the implementation of the pilot and the proof of the concept that has been developed in the first period of the project.

The technical aspects are covered by the project “Vision Infrastructure Project on Secure Infrastructure for CONFidential data access” (VIP-SICON-Project) which is conducted with Eurostat and an external company to set up a platform where the microdata is hosted to access and analyse the EU statistics. The VIP-SICON System is a framework designed for different purposes, i.e. European Group Registers (EGR), DARA and even more. The main goal was to set up a secure infrastructure for sharing and accessing EU confidential data from Member States. Within this system the confidential data can be viewed, processed and analysed. The focus is on various parts like IT infrastructure, technological aspects, security issues and procedural aspects i.e. guidelines, documentation of the infrastructure and the workflow.

Part of VIP-SICON is to provide the system and a platform where the data can be analysed inside Eurostat. Part of ESSnet DARA was to define how a remote access system will look like, how it will be implemented and how to connect to the Eurostat system.

The VIP-SICON project is conducted to build up an IT system for a remote access via thin-client and to set up the working platform and applications for researchers to analyse the data. The task of the ESSnet DARA was to connect the NSIs to the central system at the server at Eurostat.

3 Results

The ESSnet DARA project has implemented and tested a pilot of a European Remote Access System. Next to the physical installation, the project team has drafted several documents for technical descriptions, list of user and safety requirements for a European remote access system, manuals for NSI staff and researchers and investigations for alternatives.

The ESSnet DARA project had two main focusses. On the one hand it has developed the theoretical requirements like documents, manuals, user and security demands and on the other hand it has done a practical proof of concept with a real pilot implementation of a remote access system from a Safe Centre to a central node.

3.1 The pilot

Originally it was planned to establish a secure network between the Member States and Eurostat as the central node. For this reason the project “Vision Infrastructure Project on Secure Infrastructure for CONFidential data access” (VIP-SICON-Project) was set up to build a platform inside the European Commission network to access and analyse the EU statistics. At an early stage it was planned that the ESSnet DARA application of the VIP-SICON system will be built according to the list of user and security requirements for a European remote access system.

For different reasons the VIP-SICON system was not available until the end of the DARA project. Therefore the project team has started to implement an alternative pilot. In this scenario the central node (Eurostat) is represented by a Member State (France), please see Fig.1.

This alternative pilot is based on the existing French remote access system, but specially designed, converted and implemented for an European remote access system. It fulfils the demands that have been developed in the list of user and safety requirements. GENES in France provides a highly secure remote access platform, its specific procedures and the IT security. It is designed as a proof of the DARA concept, but not applicable for a real production phase. This is not because of IT aspects; it is mainly caused by legal concerns of the Member States (MS). The data sources remain in one unique central node infrastructure (isolated from the production network and internet). Part of the project is to define how a remote access system will look like and how it will be implemented.

The solution is focusing on various aspects such as IT infrastructure, technological features, security issues and procedural aspects i.e. guidelines, documentation of the infrastructure and the workflow.

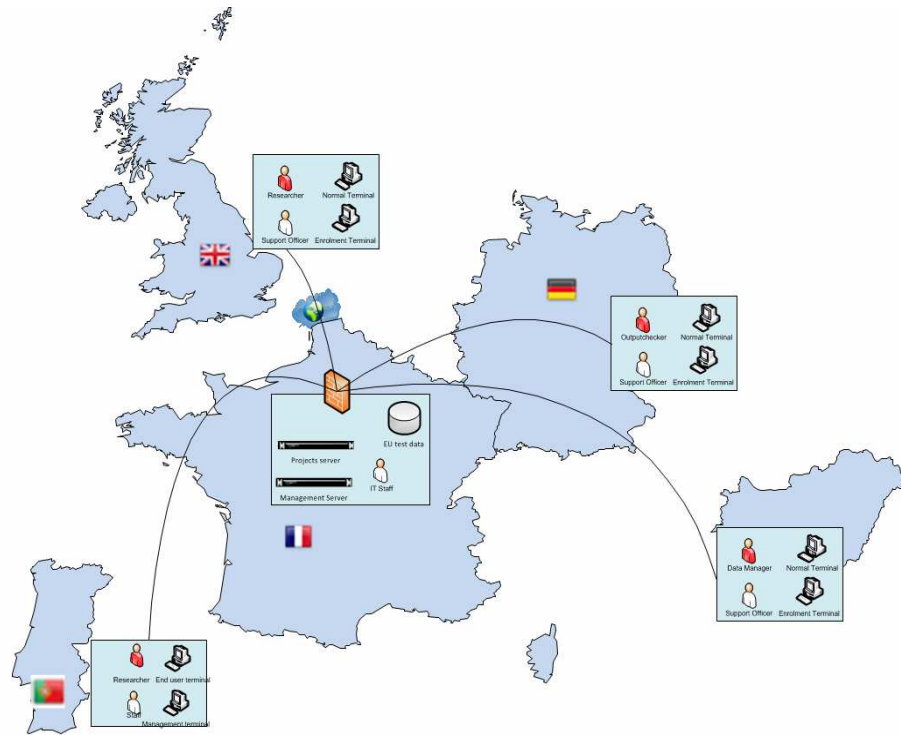


Fig 1. Model of DARA pilot; source ESSnet DARA, GENES.

The system has a strong user authentication based on certificate (PKI) to prevent any data leak, so that the user cannot get data out of the system. It shall also provide an easily access and usage for researchers with a common environment with scientific and editing software.

In the framework of working groups within the ESSnet DARA project, a functional analysis of DARA remote access centre has been made and defined five identified roles

1. **The Researcher:** has access to microdata and different statistical software within full Windows desktop. This can be extended with the usual packages, modules, ADO, etc. In order to complete the work, a researcher may have to import data into his or her workspace and request to export some of the results.
2. **The Support Officer (SO):** has the role to provide support for researchers and serves as a single contact point between the researcher and other users in the process. He has a central role in the communication with the researcher and the other persons involved (one face to the costumer).
3. **The Output Checker (OC):** the main responsibility of the Output Checker is to check the researcher output against disclosure. It is legal requirement and also an expectation from the National Statistical Systems of the Member States to check all researcher outputs against disclosure before they are made available to the researchers for further use.

4. **The Data Manager (DM):** The Data Manager is responsible for preparing the researcher datasets and metadata. He or she transfers them directly into researchers' workspaces. The researchers never get in touch with the Data Manager directly; it is the role of the Support Officer to act as an intermediary between the researcher and the Data Manager.
5. **The IT Manager:** The IT Manager is responsible for the deployment, set up and maintenance of the remote access system. The IT Manager never gets in touch with researchers directly.

One person may take two or more roles; sometimes several people share the work of one role. The remote access system is designed to consider the different needs and rights of the single roles.

The main advantage of this remote access system is that no data and not even anonymised data will physically be transferred to the MS. The data will remain at the safe servers at the central node and will be only accessed via secure connection from Safe Centres in NSIs which are also in a secure environment of official statistics. From data owner perspective this system generates a lot of trust because a decentralised data management with local copies will be avoided and Eurostat is able to keep control of the data.

3.2 Documentation and manuals

The results are deliverables with descriptions of the remote access system inclusive the workflow, a real technical infrastructure and also valuable experience that has been made during project and the testing phase of the pilot.

The ESSnet DARA project has developed a comprehensive **“list of user and safety requirements for an European remote access system”**. This paper describes the details of user requirements, the safety requirements from IT perspective and also the workflow that involves all parties to maintain the system. This list supports the VIP-SICON-Project to design the working platform for the researchers. A cost benefit analysis has been made as well to estimate the costs of a real implementation and maintenance of a remote access system inclusive manpower involved.

All final deliverables that have been produced refer to the new regulation (EC) No. 557/2013 and its “Guidelines for the assessment of research entities, research proposals and access facilities”.

During the project there have been drafted documents with descriptions and guidelines for NSI staff and researchers with the following content:

- **Accreditation Guidelines for Safe Centres**
- **Technical description of DARA pilot**
- **Review process for research proposals**

- **Process and templates for signing confidentiality undertakings**
- **Researcher management system**
- **Manage user accounts**
- **Data management**
- **Output Checking**
- **User manual on access to confidential data**

All the documents are combined to a single handbook which is for the use at the EC, the NSIs and a separate one with essential information for the researchers.

Another EU project, *Data without Boundaries* (DwB), supports equal and easy access to official microdata for the European Research Area aiming researchers' transnational access to official microdata through coordination of existing infrastructures. Both projects aim to develop compatible solutions for remote access in Europe. The best solution is, if the system is synchronised and transparent but leaves also room for flexibility to integrate other technologies. Therefore an exchange and a lively communication between the two projects had been taken place.

4 Outlook

The DARA pilot shows, that a secure, user friendly and affordable system for a European microdata access is possible in practice. The project has made valuable experiences and has produced results, which are applicable for the whole ESS. Nevertheless, there are still many of tasks remained. The central node needs to be implemented inside the European Commission's network to generate trust between Member States and gain experiences out of a running system. For a real production phase the Member States need to agree on such a solution. The Member States also have to make minor investments to implement the system, therefore support and expertise must be guaranteed for the expansion of the network. It is also recommendable to start with a few Member States for a real production phase and involve more and more Member States successively.

A lot of investments have been made already and in the long-term, remote access seems to be the best option (Brandt and Zwick, 2011; Brandt and Eilsberger 2010; ESSnet DA). The technical developments have reached a phase where online access is possible from anywhere or will be possible soon with the relevant range.

Remote access allows researchers to process the data at any time and location and has the advantage that the data remain in the protected environment of the statistical offices. Also, that kind of data access increases the networking among researchers and the scientific transparency because any researcher may access the data and replicate results any time.

The possibilities offered by the technical infrastructure are far from exhausted and have further potential for development in the future. Nevertheless legal issues, too, must be settled at this point.

The aim of the project is, to provide the basis for remote access. The project has taken into account the methods that have been developed in other countries and has benefited from the experience of international working groups.

References

- Brandt, M. et al., (2011): ESSnet Multi-Beneficiaries Grant Application: *ESSnet project ESSnet on Decentralised and Remote Access to Confidential Data in the ESS*, Luxembourg.
- Brandt, Maurice und Markus Zwick, (2011): *Decentralised Access to Confidential Microdata in Europe*, New Techniques and Technologies for Statistics, Brussels.
- Brandt, Maurice und Patricia Eilsberger, (2010): *The ESSnet-Project “Decentralised Access to EU-Microdatasets”*, Joint UNECE/Eurostat work session on statistical data confidentiality, Bilbao.
- Croessmann, Anja, (2011): *“ESSnet on Decentralized and Remote Access to Confidential Data in the ESS”*, Joint UNECE/Eurostat work session on statistical data confidentiality, Tarragona.
- ESSnet *“Decentralised Access to EU-Microdata Sets”* (DA):
<http://www.cros-portal.eu/content/da-finished>
- ESSnet *“Decentralised and Remote Access to Confidential Data in the ESS (DARA)”*:
<http://www.cros-portal.eu/content/decentralised-and-remote-access-confidential-data-ess-dara>
- VIP-SICON-Project: <http://www.cros-portal.eu/content/vip-sicon>
- DwB project website: <http://www.dwbproject.org/>