

**WP. 34**  
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 (vii): Trans-border access to microdata

## **ESSnet on Decentralized and Remote Access to Confidential Data in the ESS**

Prepared by Anja Crößmann, Research Data Centre of the Federal Statistical Office, Germany

# ESSnet on Decentralised and Remote Access to Confidential Data in the ESS

Anja Crößmann\*

\* Research Data Centre of the Federal Statistical Office Germany, e-mail: anja.croessmann@destatis.de

**Abstract:** One of the major tasks of the European Statistical System (ESS) is to provide reliable information about the society, economy, environment and development in the European Union. The access to these confidential data is available by sets of anonymised microdata or in the premises of Eurostat in the safe centre in Luxembourg. To improve the microdata access for researchers, the project “ESSnet on Decentralised and Remote Access to Confidential Data in the ESS” deals with the implementation of an access to confidential data in Eurostat through a remote desktop connection in the safe centres of the National Statistical Institutes (NSI).

## 1 Introduction

By now the European Community authority may grant access to twelve surveys for research purposes. The access to these confidential data can be provided by sets of anonymised microdata or in the premises of Eurostat’s safe centre in Luxembourg. Because of the changing demands of user needs the request for microdata tends to the usage of original microdata without direct identifiers. The anonymised microdata are a valuable source and the 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. For this purpose there is currently only the possibility for European researchers to travel to Luxembourg to get access to the data in the safe centre of Eurostat.

One way to guarantee decentralised access for researchers to Community Statistics could be to provide access in the National Statistical Institutes (NSI) of their Member States. The ESSnet project “Decentralised Access to EU Microdata Sets” has proven the feasibility to access confidential data from Eurostat through a remote desktop connection in the safe centres of the Member States. As a central result of this project the recommendation for the implementation of a pilot study was given. The project “ESSnet on Decentralised and Remote Access to Confidential Data in the ESS” deals with the implementation of the remote access from safe centres in Member States (MS) to the community statistics from Eurostat exemplarily for the European Community Household Panel (ECHP) data.

The ESSnet project on “Decentralised and Remote Access to Confidential Data in the ESS” is divided into two parts. The first part is an administrative, organisational and preparing phase. The second one is an implementation phase where the connection of a safe centre via remote access to the remote access system in Eurostat (VIP) shall be executed. The duration of the action is planned within 24 months.

The administrative part is an essential preparation phase which will also bridge the time gap until the remote access system is up and running at Eurostat. In the second phase the project is depending on the infrastructure to connect the safe centres to the central node in Eurostat. If the infrastructure at Eurostat is not available on time, one alternative could be to draw on an existing NSI’s remote access to solve this issue.

## **2 Documentation and workflow**

For a practical implementation of a remote access system from Member States to Eurostat a modification or revision of the legislation is necessary. Legally it is only allowed to provide access to community statistics in the premises of Eurostat and via anonymised microdata files. The option for remote access from a safe centre of a Member State to Eurostat is not yet foreseen. Therefore the working group on statistical confidentiality (WGSC) has convened a taskforce to assemble the revision of the EC 831/ 2002 regulation, after which remote access options from safe centres could be implemented. To access European community statistics in a safe centre of a NSI via remote access it is essential to have a thought-out documentation for the employees of the NSIs as well as for the researchers about how to get access to these data. Also a workflow has to be developed, which contains a scheme of which steps need to be undertaken and who needs to be involved to provide access. Based on the recommendations in the final report of the project “Decentralised Access to EU Microdata Sets” the tasks for the application process shall be considered.

Thereby it is necessary to define rules and standards for the data access. This includes the definition and coordination of who is permitted to get access to European microdata. The current system has to be reviewed on behalf of potential necessary modifications. A proper way to accredit scientific researchers has to be determined. One subtask will be to compose a guideline on how to communicate the possibility of the new data access option. Common standards of admissibility and the admission procedure have to be negotiated among all Member States.

Furthermore it is planned to compile a manual for the scientific community on how to get access to European community statistics within a safe centre, including a checklist for prospective users on how to request access to European microdata. Decision processes have to be communicated so that they become a transparent and fair procedure.

A standardised request form for data access has to be drafted. As all Member States have to agree that their data will be used, a thoroughly elaborated request form is necessary that compiles all relevant information on research proposals in a transparent and readily comprehensible way. The request form can be made available on the projects homepage and other relevant homepages.

Additionally a manual for the NSI regarding the workflow when data access is requested is required. The manual should describe the chores of the local RDC which provides advice to the researcher seeking data access as well as establish and describe the consultation process among all affected MS and Eurostat.

Access to European microdata can only be successful if all NSIs will be convinced that their data will be accessed in a safe and sensible way. Therefore, the workflow of processing research projects and the decision making along the way have to be well documented and communicated. On the other hand, the acceptance of this new form of data access among the scientific community depends on the support they receive when applying for admission. A fast and clear admission process contributes to a higher rate of acceptance. Another key element is the development of a common procedure for securing statistical confidentiality.

### **3 Concept of technical implementation and safety requirements**

Before the technical implementation of remote access will be executed it has to be assured that the safety requirements will be fulfilled. A safe environment for the data was already defined within the accreditation system for safe centres. In this case the concept of technical implementation usually needs to be supported by the IT division and the data security officer. There is need for a detailed description of all settings for the remote access system. From national perspective, on the one hand, it has to be assured that it is not possible to take any physical data out of the NSI. On the other hand it has to be guaranteed that nobody but the authorised researchers get access to the community statistics - not even the employees of the NSIs. To tackle this issue in an adequate way, a system of risk management could be helpful. This shall also be addressed in this WP.

For this WP it is advisable to get in touch with and synchronise the work in the VIP-Project in Eurostat and the European Group Register (EGR) which is also going to connect to the remote access platform in Eurostat to assure that the interfaces will be working between possibly different systems. Of course, in this WP, all the feedbacks from the Eurostat RDC project will be considered as the technical environment in which the solutions should be deployed. And vice-versa, all the intermediary specifications of this WP will be sent to the EGR-VIP projects. For a good collaboration, at least one mid-term meeting (t+5) will be organised.

The results and recommendation in the final report of the project “Decentralised Access to EU Microdata Sets” shall be considered here as well. This is the central WP which defines how a remote access system will look like and how it will be implemented. Therefore it is important to take the following aspects into account.

Eurostat builds an IT system for a remote access via thin-client; it is envisaged that the NSIs would be able to manage their own ‘areas’ of the central system, setup and manage accounts for researchers. For the NSIs it is necessary to have common methods to access the central system.

For the data preparation it is conceivable that each country manages its own national area via remote access inside the European central data storage (imagine 27 different folders, each with national microdata sets). Each MS gives - under its own decision - its own data to the researcher, by copying them to the corresponding user data area.

One very important issue in the remote access to European Community Statistics is the clearance of the output. It is envisaged that the local RDC checks the output, making use of European guidelines, and makes decisions on its own authority. Researchers get the results emailed to them and a copy is kept at the RDC. A system for peer review will be set up. It would be composed of one representative of at least three partners. Each member should have knowledge of output-checking (for instance person in charge of output checking in his or her country). Eurostat would be associated in this peer review system. The main objective is to analyze the output that has to be checked and to define the best way to ensure trust among NSIs in the output checking process. Some proportion of output can be double checked by another NSI or Eurostat via a specific workflow. One important thing is that all the outputs should be stored in the IT-System for analyzing and reviewing.

A central simple administrative system should be made available. Local RDC personnel can remotely log on to this system to add some basis information on the new contract (name of the institute, research aim, name of researchers, data sets used, start date, finishing date, etc). This central administrative system could be placed on the central IT-system at Eurostat, for instance on a secure website.

#### **4 Cost benefit analysis**

A cost benefit analysis gives an overview about the costs of the implementation so it is possible to estimate what an up and running network will cost. Based on the experiences of already existing RDCs it should be possible to calculate the costs of the hardware that allows the access to microdata either on a national server or via remote access. But the implementation of new ways of accessing community data surely leads to an increasing demand that causes additional staff as well. Also for NSIs that are aiming to implement a RDC, an estimation of the occurring costs should be useful.

On this basis, a cost model that gives information on the staff unit costs, the scale of operations, the operating costs per project and finally the share of costs split by categories can be estimated. Also with regard to further implementation projects it is necessary to discuss on how the financial burden will be covered and which costs are eligible. This decision is still remaining even if Eurostat is willing to support the costs or if the NSIs are required to self-finance the service. If last-mentioned occurs, a (partial) assumption of costs could be covered by the users, for example the costs for the hardware and set up of the thin client.

Based on the cost template that has been developed in the former project, the different categories shall be included and evaluated on a real implementation. What is essential here is the staff planning (rating each qualification/grade), the breakdown on strategy and operational costs, the breakdown on fixed and variable costs, the number of projects as well as the IT costs.

The cost model shall be practically applied and implemented for all partners who will implement the pilot remote access system to have a real cost calculation for future cost planning. However, it is also important to take the benefit for the European research community into account. If there is no well functioning data infrastructure, the community statistics can't be used in a proper way. The national experience shows that the data will be heavily used if the infrastructure, with an administrative burden as low and also as safe as possible, is established.

## **5 Implementation of remote access – case study**

After the conception of the documentation and the workflow, the technical and safety requirements and the cost benefit analysis, now a real implementation of a remote access connection from a safe centre in an NSI to the remote access platform in Eurostat shall be converted. Following the guidelines for the precedent defined data the connection from a thin client PC shall be set up. The objective is to implement a pilot in real conditions from a NSI. This covers the definition of the perimeter of the pilot: selecting the data, the authorisations, the users, the client configuration, etc.

The end-point, the thin client PC, is the most vulnerable element of the connexion channel. Because it is not under the control of the IT administrators or the data-producers (in case of decentralised access), it is important that the security level is considered seriously.

For evaluation purposes, ten thin client PCs will be designed and configured with high level security constraints. The configuration will include: the PC, the monitor, a bio smart card reader, smart card (one per user), OS license, and scientific license software if needed. The final configuration will depend on the results of the defined requirements and on the final Eurostat research data centre infrastructure. It is planned to install at least two thin client PCs per NSI (one for the researcher, one for

the data-manager). All thin client PCc will be configured in one place for easy deployment and for security reasons. The users should work on the pilot at least six months.

If the remote access infrastructure of Eurostat is ready on time, the implementation from a NSI will be plugged on. Otherwise, a fallback scenario is to experiment all the process and workflows defined on the concept of the technical and safety requirements from an existing NSI's remote access. If a connection between a safe centre in a MS and the central node in Eurostat shall be established, hardware equipment in the NSI of the MS is necessary. This includes the IT hardware costs of the thin client, the set up of the safety requirements as far as not already in place.

Additionally, a set up of the IT environment to meet the security guidelines in each NSI is necessary. The procedure for the security-check will be defined in a manual of all the required specifications for security, in agreement with all the NSIs concerned. The responsible IT department needs time to test the security if the thin client system is connected with the NSI IT-infrastructure.

The main objective of this security-check is to know if it is possible to get data or data files from the system (via network data-transfer, hard-drive analysis, usb key hack, key logger, hook on the copy-past process, etc.). Moreover, it is necessary to test the impersonation of the accessing persons and to develop an authorisation procedure.

## **6 Communication and dissemination to the ESS**

Next to the communication and dissemination of the results to the non-participating NSIs in the ESS it is necessary to promote the results to the research community to inform about the existence of a way of comfortable access to community statistics by now. It is imaginable to organise an European research award for social and economic relevant results based on European Statistics.

It is also very important to connect the project with other projects and groups of interest dealing with the task of providing and expanding access to European data sets via remote access. Main objective is to participate in the particular meetings and workshops to exchange knowledge and give feedback to the relevant circle.

The most crucial issue is the project "Design and development of infrastructure/service for remote access/execution in Eurostat" (VIP-Project), because the "ESSnet on Decentralised and Remote Access to Confidential Data in the ESS" is highly depending on the existing infrastructure at Eurostat. It is planned to synchronise the requirements and user needs for the remote access system with the project group.

Furthermore the revision of the regulation (EG) 831/2002 is in full play now to provide the legal basis for a decentralised remote access. It is very important to take the developments in the taskforce 831 into account and to give feedback from the project group to the taskforce.

One part of the envisaged project “Data Without Boundaries” (DwB) will also be dealing with the issue of providing access to and expanding the network of European Community Statistics. The co-ordinator and some co-partners are as well members of DwB. It is planned to have an active exchange and a harmonisation between the DwB work package and this project.

Besides, there is an interest group in the OECD dealing with the issue of remote access, too. The working group on Microdata Access met first in June 2010 in Paris and is also going to meet during the project time.

Likewise the developments on the “Workshop on Data Access” (WDA) need to be taken into account. This workshop is an international platform for exchanging information and knowledge between National Statistical Institutes, Research Institutions, Data Archives and Universities.

The goals, the state of affairs and the final results of the project will be presented as an overview on the ESSnet portal <http://www.essnet-portal.eu> with a link to a more detailed version on the website [www.safe-centre.eu](http://www.safe-centre.eu). The project website will contain a communication tool like a feedback forum and a contact e-mail-address, where other non-participating countries are invited to give feedback on the project and where comments are expected for example on a special legal situation in their country. An invitation e-mail with the link of this website will be sent to the responsible institutions in all MS of the EU to inform about the project and welcome them to make remarks and give input.

## **7 Conclusion**

The pilot implementation of a “Decentralised and Remote Access to Confidential Data in the ESS” exemplarily for the ECHP data will take part within two years. The project depends on different circumstances and the developments of other ongoing projects. Legal, administrative, technical and cost requirements have to be defined and pretested. If the pilot network of some few safe centres connected to the Eurostat platform is running well, a future extension to other NSIs can be attempted.

The vision of this project is to develop a common procedure which certifies safe centres in the NSIs for a decentralised and remote access to confidential data in the ESS and to connect them to the running network.



## References

Brandt, Eilsberger, Zwick (2011): *Decentralised Access to Confidential Microdata in Europe*. RDC Germany Working Paper No. 38. Wiesbaden, [http://www.forschungsdatenzentrum.de/publikationen/veroeffentlichungen/fdz\\_arbeitspapier-38.pdf](http://www.forschungsdatenzentrum.de/publikationen/veroeffentlichungen/fdz_arbeitspapier-38.pdf)

Hundepool, Anco et. al. (2010): *ESSnet Handbook on Statistical Disclosure Control. Version 1.2*, [http://neon.vb.cbs.nl/casc/%5CSDC\\_Handbook.pdf](http://neon.vb.cbs.nl/casc/%5CSDC_Handbook.pdf)