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

Safe Centre Network - Need for a Safe Centre to Enrich European Research

Prepared by Maurice Brandt, Federal Statistical Office, and David Schiller, Institute for Employment Research Germany

Safe Centre Network - Need for Safe Centre to enrich European research

Maurice Brandt* and David Schiller**

* Federal Statistical Office, Wiesbaden, Germany, e-mail: maurice.brandt@destatis.de

** Institute for Employment Research, Nuremberg, Germany, david.schiller@iab.de

Abstract. During the last years the microdata access to official statistics has considerably improved. On national level sophisticated approaches to access microdata are developed, e.g. remote access or job submission, but transnational access and international comparative research are still not easy to do. Two European projects, the ESSnet "Decentralised and Remote Access to Confidential Data in the ESS" (DARA) (www.safe-centre.eu) and "Data without Boundaries" (DwB) (www.dwbproject.org) are working on improvements of microdata access in that special area of social science research. At the moment, the common solution to work with highly detailed confidential microdata is still on-site stay. Users have to travel to the location where the data are stored and have to work within a Safe Centre. There is also the possibility to combine the advantages of Safe Centres and remote access in a single approach of a transnational secure network of Safe Centres to improve the access to international microdata. In that case, researchers don't have to travel to another country any more to analyse other national or international datasets; they can access the data from a Safe Centre in their own country. This paper stresses out the need for an organizational network of Safe Centres bound together by secure remote access connections.

1 Introduction

Scientific research is important for the development of societies. To support high quality research, access to highly detailed data on individual level (persons, households, enterprises) becomes also more and more important in an international context. Enlarging the use of such data sources goes along with high demands on disclosure control. According to that highly detailed confidential microdata is in the first place only possible via on-site stay. Users have to travel to the location where the data are stored and have to work within a Safe Centre. Technical improvements like secure remote access solutions enable access to confidential research data from a remote location via secured internet connections (*Report on the state of the art of current SC in Europe*, 2012). A crucial point for such solutions are the access devices and access locations. A number of data owners, responsible for research data with a high demand in security, need control about the devices and location used to access research data via secure remote access. To accommodate with that need the advantages of "Safe Centres" and "remote access" can be brought together in a single approach (Brandt & Eilsberger, 2009; Bender & Heining, 2011). In that case, researchers don't have to travel to the home location of the data owner; they can work within a safe centre close by or even in their own institution and access the data via a secure remote connection. This paper describes the need for an organizational network of safe centre bound together by secure remote access connections.

2 Safe Centre projects

On European level two projects, the ESSnet on "Decentralised and Remote Access to Confidential Data in the ESS" (DARA) (www.safe-centre.eu) and "Data without Boundaries" (DwB) (www.dwbproject.org) are working on improvements of microdata access in that special area of social science research. The two European projects are dealing, amongst other things, with the Safe Centre approach in combination with a secure remote access connection.

2.1 ESSnet project "Decentralised and Remote Access to Confidential Data in the ESS" (DARA)

The idea of an European network of Safe Centres has started in the year 2008 and was carried out in the ESSnet Project "Decentralised access to EU microdata sets" (DA) during the year 2009. The goal of this project was to study the feasibility of setting up a network of Safe Centres that enables access to confidential microdata sets for research purposes throughout EU countries. Various scenarios, procedures and technical solutions for sharing microdata sets amongst Safe Centres have been investigated. One task was to develop the methodology, guidelines and requirements that are compatible with the legal frameworks and practices with respect to the handling of confidential data in the European Statistical System (ESS). One goal was to support National Statistical Institutes (NSIs) to set up and to adapt their infrastructure in order to comply with minimum requirements for hosting European microdata sets.

One central result of this project was the recommendation to establish a network of Safe Centres, which are connected via remote access to a central node (Eurostat), where the data are stored on a secure server. The implementation of a pilot study had been proposed (Brandt & Zwick, 2011). For the results of the project see also the final report of the project ESSnet DA (http://www.cros-portal.eu/content/da-finished).

The implementation of the pilot was conducted from 2011 to 2013 by the follow up ESSnet project "Decentralised and Remote Access to Confidential Data in the ESS" (DARA). The six partners of the ESSnet project DARA are working on 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. The ESSnet DARA project had two main focuses. On the one hand it has developed the theoretical background and on the other hand the ESSnet DARA project has implemented and tested a real pilot of a European Remote Access System from five Safe Centres in Europe to a central node in France. 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 (Brandt, 2013, for more detailed results on this project). 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 (Eurostat) and will be only accessed remotely 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. One task was to harmonise the goals and developments with the EU project, "Data without Boundaries" (DwB). Both projects aim to develop compatible solutions for remote access in Europe.

2.2 EC FP7 project "Data without Boundaries" (DwB)

The European Commission Framework Program 7 (FP7) founded "Data without Boundaries" (DwB) project runs for four years; from May 2011 to April 2015. 29 Partners from eleven European countries work on improvements for transnational research in Europe. Within the project twelve work packages (WP) deal with topics like resource discovery, data documentation and accreditation (for more information visit www.dwbproject.org).

WP 4, termed *"Improving Access to OS microdata"*, focuses on technical developments to improve access to confidential microdata. The focus is thereby on secure Remote Access (RA), respectively on Remote Desktop solutions, defined as followed: *any kind of terminal or desktop solution that uses a secured connection to the servers of the respective data provider, whereby the user can see and work live with the real, highly-detailed and confidential microdata.* The first task of this WP resulted in deliverable 4.1 (*Report on the state of the art of current SC in Europe*, 2012). A survey, focused on CESSDA archives and European NSIs (National Statistical Institutes), with focus on the technical perspective of remote access was carried out. Eight solutions are described in the report. When looking at locations from where data is made accessible, three types were found:

1. There are "pure" remote access solutions that are not restricted to a given access point and researcher can use data via the internet from anywhere.

2. The most often used approach is allowing access from a given room within an institution (e.g. a university) The researchers contractually agrees on a number of restrictions, e.g. accessing data from this room only, locking the room so nobody else can access data. In parallel, a given IP-range (or specific IP addresses for terminals) is needed to be able to reach the distant data source.

3. Finally, highly-detailed data sources are only accessible via remote access when the connection is established from within a Safe Centre. This concept can be seen as Research Data Centre in Research Data Centre approach (Bender & Heining, 2011).

Starting from the findings of the survey the partners of WP4 proposed a European infrastructure to improve access to confidential microdata and ease transnational research (*Feasibility study on the organizational architecture for managing pan European access*, 2013). This infrastructure, the European Remote Access Network (Eu-RAN), is build around an "Single Point of Access" that is equipped with a number of services and nodes together users and user groups from different locations with data sources stored in different locations (Schiller, 2013). Safe Centres are one access point supported by the proposed infrastructure.

3 A Safe Centre

Remote Access understood as remote desktop (or terminal) solution differs from job submission solutions. Whereby in the first case the researcher can really browse the datafile and directly send enquiries and receive outputs; in the second case the dataset can not be seen and enquiries are send into a kind of black-box; the outputs are normally checked before they are send back to the researcher. In this paper we focus on remote access where the researcher can work live with the data (remote desktop). This data access approach is much more comfortable for the researcher; at the same time the screen may show more confidential information as when using job submission solutions. The crucial point is therefore the access device or in other words the screen used by the accessing researcher. Safe Centre solve the problem of not knowing exactly what happens in front of the screen whilst building a uninterrupted safeguarding of the confidential research data.

The following definitions should give a clearer understanding of the Safe Centre concept: A Safe Centre is not the same as a access facility mentioned in EC regulation 557/2013. While an access facility is an organization that is permitted to give access to EU statistics after a successful accreditation procedure, a Safe Centre is a physical and organizational structure that enables access to confidential microdata from different European parties via a secured remote access connection. This means the Safe Centre can be located inside an access facility but an access facility does not necessarily need to have a Safe Centre. A Safe Centre can be accredited and become officially an access facility for EU statistics. The other way around, a Safe Centre can exist without being an access facility, because the Safe Centre is a standardised access point that enables access to confidential microdata in general, i.e. not only EU statistics. The same applies to term Research Data Centre (RDC) and/or Remote Access Centre. An RDC, providing research data to the scientific community in a standardized manner, may function as a Safe Centre but it can also only rely on other access ways, e.g. on-site stay, remote access (without Safe Centre), job submission and/or download of anonymised files. A Remote Access Centre can offer Remote Access without using Safe Centre but it would be easy to add this access way to his portfolio. Safe Centre do not have to be located in the facilities of data providers. There are two reasons why they are often located in such facilities: first, because the data provider often have the needed rooms already available (for on-site stay), second, because the staff members of existing data providers have a very high knowledge about data security issues and are therefore trusted partners and reliable guardians of data access points. At the same time there are no reasons against other Safe Centre locations, e.g. at universities, if trained staff is available and contracts secure the safety standards.

A Safe Centre consists of an organisational part and a safe room representing the physical part of the Safe Centre concept. Safe rooms are rooms especially designed to give access to confidential microdata.This means that only devices to access the microdata are allowed in the room, no access to other information resources is possible. There is a physical access control and some kind of surveillance; trained staff is labelled that cares for the security standard and arrangements are made that make sure that only entitled persons can have access to the devices (the last point refers to things like visual covers on windows and barriers within the safe room protecting single access devices, i.e. screens). The exact design of a Safe Room depends on the security needs of the data providers. The organisational part of the Safe Centre guarantees a secure procedure of data access by carrying for measures like contracts, agreed on work flow, responsibility and accordance with security restrictions. This all needs cooperation leading to agreed standards for access to confidential microdata in Europe. One additional result of such cooperation activities is the setup of trust between the involved partners.

There are a number of barriers persons needs to pass before they get results from out of the system. Apart from the technical barriers, there are also barriers that are secured through the work flow of granting access to confidential data.

1. First of all a contract, fixing all relevant measures needs to be signed between the host of the Safe Centre and the data owner.

2. Second of all a contract needs to be signed that qualifies a certain researcher to access the data.

3. There are physically identification checks at the entrance of a Safe Centre.

4. Login information are needed for the access device in the Safe Centre as well as for the secured remote connection to the servers of the data provider.

5. The server at the central node requires an authentication as well.

6. And of course there is always the output checking procedure at the end, which assures that only safe output is released to the researchers from the secured remote access environment.

The mixture of standardised physical and organisational measures to enable comfortable and secure access to confidential microdata shows the concept behind the term Safe Centre.

4 Pros and Cons of a Safe Centre

Having such a Safe Centre in place results in a number of advantages for both sides; researchers and data providers:

- Less travelling is needed to access confidential microdata. This advantage should not be underestimated. Just one example: Researchers living in Portugal have to invest huge efforts if they want to work with data from the UK. In general a on-site stay for a couple of weeks is needed. With the possibility to use a Safe Centre in Portugal to access the UK-data research becomes much more efficient. The researcher can work on the project for shorter time frames and can easily go back to his office between those time frames. Also for shorter distances a Safe Centre nearby will definitely ease transnational research in Europe.

- Data providers can fulfil there duty to make data accessible for scientific research without an increased disclosure risk. Data can be made usable in a distant location without losing control about the security measures.

- More and more data sources are available for scientific research. This big advantage for researchers goes hand in hand with the drawback of more and more complicated anonymisation procedures. Detailed information and high sample sizes boost the disclosure risk. As a result it is very likely that more and more data sources are only available after rigorous anonymisations and with less usability for researchers. Safe Centre can ease this future challenge by offering a secure way to make highly detailed and confidential microdata available for scientific purposes.

- On an international level establishing such Safe Centre will result in an institutionalisation of transnational microdata access.

Nevertheless there are also some disadvantages:

- Even if only shorter travel distances are needed to have access to confidential microdata; travelling is still needed.
- Effort is needed in order to set up contracts and agreements.
- Costs for setting up and maintaining a Safe Centre occur.
- Trained staff is needed and has to be paid.

It is important to note that Safe Centre are only one approach within the data access portfolio. Scientific use files, remote access solutions without the need for a Safe Centre, on-site stay and job submission have still their right to exist and more than that: they are needed to make the whole range of available research data accessible for scientific research. What access way might be the best, depends of the kind of data, the degree of anonymisation and the specific research project.

A Safe Centre might be not the first choice for a modern national microdata access but a network of Safe Centre for international microdata access is a real improvement, because the distances for travels are much shorter if the Safe Centre is located in the same country. If the Safe Centre is located in the same city like university and other institutions, there are almost no travel costs at all. Some European countries are using remote access and Safe Centre parallel. Many European countries do not have remote access. Those countries might allow access to their data in the controlled environment of a Safe Centre but not via remote access connection to the research institutions.

5 Need for Cooperation

The exact design of a Safe Room depends on the security needs of the data providers.

A Safe Centre Network will only be useful for the researchers, if it is build on common standards. Cooperation in order to find standards for Safe Centre and Safe Rooms that allow to build a Safe Centre Network for Europe (and maybe beyond) are therefore most important. Every project that deals with a Safe Centre approach should be aware of other projects and try to coordinate the actions undertaken. Thereby coordination is needed on different levels; namely on organisational, technical and legal level.

The ESSnet DARA project is embedded in a clear framework. The main differences to the DwB project are, that DARA refers only to EU statistics under the regulation (EU) 557/2013 (community statistics). It provides a microdata access solution for the ESS, starting with NSIs only. After (good) experiences after a few years, the ESSnet DARA system might be expanded to other data producers and access points. The DwB project unifies NSIs, data archives and universities. The proposed European Remote Access Network (Eu-RAN) should act as a service provider for all data providers in Europe. According to this goal, solutions for all potential users of Eu-RAN have to be found. Regarding data providers that need a Safe Centre to allow access to their data intensive coordination work is needed to find a common solution.

5.1 Organizational cooperation

In a network it is essential that the different institutions that are involved communicate and cooperate. Within a Safe Centre Network a common organisational framework is needed. This framework has to care about security requirements, responsibilities, operation cycles, etc. Discussing such things and finding suitable solutions will establish trust between the partners; knowing that the others care for the data in the same way as the producer does.

5.2 Technical cooperation

For a common technical infrastructure for a remote access system special standards for internet security like encryption standards of connection and security standards for the data server need to be agreed. Only if the authentication procedure, the level of security for the connection itself (like VPN settings) and the configuration of the endpoint is considered as safe enough to provide access, data producer are willing to join and allowed to do so. Beside the agreed security standards, establishing and maintaining a network is a crucial topic. The pros and cons of centralised and de-centralised approaches have to be discussed.

5.3 Legal cooperation

First a evaluation of legal restrictions with respect to different European countries and different kinds of research data is needed. The findings have to be brought together and the question, if the legal framework meets the needs of a Safe Centre Network has to be answered. Thereby the setting of this network - moderating between data usability and data security - has to be brought together with the needs of the legal frameworks in Europe. Always keeping the goal of reaching the best possible solution to support both aspects in mind.

6 A Safe Centre Network

There are different levels of an Safe Centre Network in Europe. From the organisational level the network builds common standards and trust regarding the shape of a Safe Centre; this refers to topics like equipment of a safe room, how should the access control and monitoring of researchers be carried out, ect. Also part of the organisational level are agreements regarding work flow and responsibilities. On a geographical level the Safe Centre Network should ease travelling for researchers by covering the relevant locations in Europe. One approach could be to build a network of Safe Centres where researchers do not have to travel more than 300 kilometres; another approach could locate Safe Centres in areas with a high populousness. From the technical level the network has to connect the Safe Centre with the research data sources in Europe. First important findings derive from the ESSnet DARA project; DwB provides the proposal for a European Remote Access Network to support the idea of a Safe Centre Network. Thereby the question of centralised or de-centralised solutions (or something in between) is crucial.

To finally reach the goal of a running Safe Centre Network in Europe a lot of efforts have to be made. On the other hand the advantages of such a network are enormous. Highly detailed and confidential data contains rich information for modern scientific research. A Safe Centre Network will enable the use of such data from all over Europe for researchers from all over Europe. In addition a infrastructure will be build that handle future data sources - including big data - as well.

7 Summary and Outlook

There are good remote access solutions in several countries that are working quite well since years by now. From a technical and security point of view, the remote access systems can be considered as safe. But from a legal point of view it is not allowed for all the countries to run remote access solutions. This heterogeneity is difficult when it comes to cross-border or international microdata access. The data providers expect that their data will be accessed at least in the same secure manner like in their own country. An accredited Safe Centre can help to assure those data providers that their data are accessed within a secure endpoint that fulfils certain requirements and not from anywhere. The final goal is to move the endpoint to the researchers institution so that they can work in their familiar environment. Due to more complex, large and sensitive data that can't be transferred to another location, it is even more important to give access from a distant access point. At the same time the demand for data confidentiality is increasing, so that it is essential to oversee the complete process of data access. Therefore a controlled endpoint is needed and this means that Safe Centres might be also indispensable in future.

References

Bender, S., & Heining, J. (2011, Fall). The Research-Data-Centre in Research-Data-Centre Approach: A First Step Towards Decentralised International Data Sharing. *IASSIST Quarterly*, *35*(3), 10-16.

Brandt, M. (2013). Improvements of access to European microdata. In *Joint UNECE/Eurostat work session on statistical data confidentiality, Ottawa.*

Brandt, M., & Eilsberger, P. (2009). The ESSnet-Project Decentralised Access to EU-Microdatasets. In *Joint UNECE/Eurostat work session on statistical data confidentiality, Bilbao.*

Brandt, M., & Zwick, M. (2011). Decentralised Access to Confidential Microdata in Europe. In *New Techniques and Technologies for Statistics, Brussels.*

Feasibility study on the organizational architecture for managing pan European access. (2013, July). www.dwbproject.org/deliverables.

Report on the state of the art of current SC in Europe. (2012, September). www.dwbproject.org/deliverables.

Schiller, D. (2013). Proposal for a European Remote Access Network (Eu-RAN) - main components. In *Joint UNECE/Eurostat work session on statistical data confidentiality, Ottawa.*