Joint Safe Centre of the Hungarian Central Statistical Office and the Hungarian Academy of Sciences

Beáta Dobáñyné Nagy, Eszter Nagy, Zoltán Vereczkei (Hungarian Central Statistical Office), János Köllő, Endre Szabó, Melinda Tir (Hungarian Academy of Sciences) (Hungary)

beata.nagy@ksh.hu, eszter.nagy@ksh.hu, zoltan.vereczkei@ksh.hu, kollo.janos@ktk.mta.hu, szabo.endre@ktk.mta.hu, tirl.melinda@ktk.mta.hu

Abstract and Paper

The Safe Centre access for scientific purposes is the most prominent among the six channels that the Hungarian Central Statistical Office (HCSO) has put at the disposal of users to obtain access to data. The rules for this type of data access are stipulated in accordance with Act CLV of 2016 on Official Statistics and Commission Regulation (EU) No 557/2013 on access to confidential data for scientific purposes. The Safe Centre access is the main accession channel available to researchers who would like to conduct their investigations using microdata collected or handled for statistical purposes by HCSO. The institutional context for the Safe Centre is the HCSO’s aim to ensure as wide a range of access as possible to datasets managed for purposes of official statistics, in consonance with the HCSO strategy of strengthening relations with other institutions, particularly scientific. The HCSO has put the development of the Safe Centre access at the forefront of its priorities due to the strategic importance and usage of the data access channels for scientific purposes. The modernisation of data access was in part boosted by good practices in cooperation between official statistics and the scientific community at the European level. The Hungarian Academy of Sciences (HAcS) has been an important partner of the HCSO throughout the years. Since the opening of the modernised Safe Centre in the premises of the HCSO in 2014, the researchers of the Academy have been its most frequent users. As a result of this partnership, a new demand has grown from the HAcS in harmony with the HCSO’s strategy: the establishment of a new, jointly-operated, remote-controlled Safe Centre in the premises of the HAcS. The paper describes from different viewpoints the set-up of the new remote-controlled Safe Centre, exploring the legal, informational technology and organisational aspects in detail. The process of data access in the new Safe Centre, the roles of the HAcS and the HCSO are also discussed. Furthermore, the paper introduces the advantages of the Safe Centre, as seen from both the official statistics and research community perspectives, represented by the HCSO and the HAcS, respectively and the cooperation between an NSI and the scientific community. It also inculdes the experiences of the past year of operation as well as data on the utilization of the facility and the datasets available.
Abstract: The Safe Centre access for scientific purposes is the most prominent among the six channels that the Hungarian Central Statistical Office (HCSO) has put at the disposal of users to obtain access to data. The rules for this type of data access are stipulated in accordance with Act CLV of 2016 on Official Statistics and Commission Regulation (EU) No 557/2013 on access to confidential data for scientific purposes. The Safe Centre access is the main accession channel available to researchers who would like to conduct their investigations using microdata collected or handled for statistical purposes by the HCSO. The institutional context for the Safe Centre is the HCSO’s aim to ensure as wide a range of access as possible to datasets managed for purposes of official statistics, in consonance with the HCSO strategy of strengthening relations with other institutions, particularly scientific ones.

The HCSO has put the development of the Safe Centre access at the forefront of its priorities due to the strategic importance and usage of the data access channels for scientific purposes. The modernisation of data access was in part boosted by good practices in cooperation between official statistics and the scientific community at the European level.

The Hungarian Academy of Sciences (HAcS) has been an important partner of the HCSO throughout the years. Since the opening of the modernised Safe Centre in the premises of the HCSO in 2014, the researchers of the Academy have been its most frequent users.

As a result of this partnership, a new demand has grown from the HAcS in harmony with the HCSO’s strategy: the establishment of a new, jointly-operated, remote-controlled Safe Centre in the premises of the HAcS.

The paper describes from different viewpoints the set-up of the new remote-controlled Safe Centre, exploring the legal, informational technology and organisational aspects in detail. The process of data access in the new Safe Centre, the roles of the HAcS and the HCSO are also discussed. Furthermore, the paper introduces the advantages of the Safe Centre, as seen from both the official statistics and research community perspectives, represented by the HCSO and the HAcS, respectively and the cooperation between an NSI and the scientific community. It also includes the experiences of the past year of operation as well as data on the utilization of the facility and the datasets available.

1. Strategy

Satisfying user needs and finding better ways to facilitate the use of the rich data assets of the National Statistical Institutes is a priority objective of official statistics all around the world. The Hungarian Central Statistical Office (HCSO) is of no exception; the Hungarian NSI defines this as one of its most important strategic goals.

The development of integrated, service-oriented systems of data access is a backbone of our strategic objectives. Therefore “we develop organisational, methodological and IT solutions, which, along with uniform practices, allow to take into account the different characteristics of each user group in the course of data releases, thus ensuring the operation of data access channels functioning according to the principle of optimal risk management and differentiated
by data providers and users and their demands, as well as the quick and effective data access through these channels". [HCSO Strategy 2020]

The needs of researchers – and the scientific community in general – is very important for official statistics. The HCSO has been investing money and energy into developing data access protocols to meet the researcher needs while at the same time guaranteeing statistical confidentiality. The researchers usually have several remarks on what they expect from official statistics to ease the burden and simplify procedures to gain access to statistical data for scientific purposes, especially microdata.

Based on our experiences, these are great opportunities for cooperation and further development of data access procedures. This need has been quite explicit both on European and national level. Thankfully to such international initiatives as the Data without Boundaries1 and the European Data Access Forums2, the data access developments were boosted by good practices and ideas from NSIs all around Europe.

In parallel to the European initiatives, the partnership between the HCSO and the Hungarian Academy of Sciences (HAcS) became much stronger. The HAcS is one of the most important users of the statistical microdata of the HCSO and good opportunities arose for stronger cooperation. The most important outcome of this good partnership materialized in a creation of the new joint Safe Centre of the HCSO and the HAcS, on the premises of the Academy. Not only it was a major step ahead in cooperation between researchers and official statistics but it is one of the most important achievements of the HCSO in recent years that brings us closer to the completion of our strategy and opening new doors for future developments on data access.

The HCSO and the HAcS proudly present the concept of this new Safe Centre.

2. Legal environment


In order to achieve this objective, Article 23 of Regulation 223 makes it possible to access confidential data which only allow for indirect identification of the statistical units may be granted to researchers carrying out statistical analyses for scientific purposes by the

1 Data without Boundaries (DwB): http://www.dwbproject.org/
Commission (Eurostat) or by the National Statistical Institutes (NSIs) or other national authorities, within their respective spheres of competence.

The above cited possibility provides the legal basis for giving access to statistical microdata for scientific purposes not only on a European level, but also on a national scale by the NSIs.

The general rules are naturally elaborated in more detail in an implementing Regulation issued by the Commission that contains the means of implementing acts, the arrangements, rules and conditions for access at Union level, however they can be regarded as guidelines for national practices as well.

Taking into account the European legal environment, the new Hungarian Act on statistics (hereinafter referred to as: Statistics Act) which has been adopted in 2016 has for the first time introduced the term microdata and access to microdata for scientific purposes. In addition, the HCSO’s internal regulation on data access defines the operating rules for the Safe Centre within the HCSO.

Before the adoption of the new Statistics Act, the access has been possible but had to be allowed based on an interpretation of the erstwhile legislation without any clear legislative background.

The legal boundaries to establish such a solution must always remain within the confines of the above regulation. These boundaries are:

- The ownership of the data must remain with the HCSO.
- The control of the safe environment must remain with the HCSO.
- The control of giving access to a specific researcher for a give project must remain with the HCSO.
- Appropriate sanctions have to be put into place in case of breach of the rules of access.

Beside the legislative background, a contract has also been signed between the HCSO and HAcS, which specifies the rights and obligations of the Parties while jointly operating the Safe Centre and establishing the financial framework for it. It contains annexes that are used as protocol for the everyday workings of the Safe Centre with special regard to safety measures.

The contract furthermore sets up a set of sanctions that differentiate between various kind of misdemeanor on the side of the researchers. The difference among the sanctions is based on the risk that the disregard for the rules represents to the security of the data. There are moreover sanctions for contractual breaches as well which eventually may result in termination of the contract.

3. **Background**

The researchers of the Centre for Economic and Regional Studies of the Hungarian Academy of Sciences have been using the Safe Centre of the HCSO since 2014 and its anonymized microdata for even a longer period of time. Since these researchers have represented the majority of the Safe Centre’s users, and as they have always respected the rules of the Safe Centre, the HCSO has been very open to the needs of the HAcS’s researchers from the beginning.
The fruitful cooperation between the two institutions throughout the years established the trust between them and resulted in a joint pledge to collaborate for many more years. In this situation, a need to open a Safe Centre outside the premises of the HCSO was emerged from the HAcS’s side. The new building of the HAcS was a perfect location for this new Safe Centre.

An agreement for cooperation was signed in 2015, which captured that the two institutions would cooperate in the process of the new Safe Centre’s establishment. The reconciliation of the legal, information technology and process organization part of the new Safe Centre began in 2016. With shorter and longer breaks, the Safe Centre in the HAcS was opened in 2018.

4. Data access in the Safe Centre

What is the process of microdata access in the Safe Centre?

The process is based on the rules of other data access channels in the HCSO that have been applied for many years. In the following paragraph, the in-depth description of the process is introduced:

- Access to the Safe Centre can be requested by the completion of the data request form for Safe Centre access.
- Safe Centre access is available only for scientific purposes and access is granted only for approved research projects that meet all researcher accreditation criteria. In addition, the HCSO needs to check whether the researcher request for microdata can be satisfied.
- Contract and confidentiality commitments are to be signed for data requests successfully approved after researcher accreditation and evaluation of the request form.
- After singing the contracts, the HCSO prepares the requested data sets with its methodological and metadata description.
- The aforementioned microdata files are then added to the safe interface dedicated to the researchers (i.e. only the researchers indicated in the request form have access to the microdata sets).

The actual research can only begin after all the above mentioned steps are completed. The researcher comes to the Safe Centre, conducts his research and prepares his outputs. These outputs are checked by the colleagues of the HCSO during the course of output checking, in order to guarantee that no disclosive data are taken out from the premises of the HCSO.

The following flow chart displays the process described above:

---

3 We would hereby like to thanks for the devoted and persistent work of the following colleagues who took a significant part of the establishment of the new Safe Centre: Zsolt Németh, Péter Soponyai, Szabó LCS László, József Padányi, Linda Andrejcsik, Edit Márder.
The process of data access in the new Safe Centre is almost the same as the previously described process, only a couple additions were made.

Firstly, the HAcS set up a professional body, the main task of which is to pre-filter the research request before the submission to the HCSO based on specific viewpoints. This becomes step zero in the process.

Secondly, the colleagues of the Data Bank of the HAcS’s are involved in the third and fifth steps of the above displayed process. Their job is to coordinate and support the data access in the Safe Centre, i.e. their role is very similar to the colleagues of the Dissemination Directorate of the HCSO. Their tasks are introduced more precisely in Chapter 6.

The following chart shows the extended data access flow in the new Safe Centre:

Who is entitled to use the new Safe Centre?

Safe Centre access is available only for scientific purposes and access is granted only for approved research projects that meet all researcher accreditation criteria, furthermore project leaders have to be in a legal relationship with the Centre for Economic and Regional Studies HAcS or Centre for Social Sciences of the HAcS. The project leader bears full responsibility for project participants entering the Safe Centre (researchers, co-authors, students and assistants).

What kind of data are available for researchers in the new Safe Centre?
Access in the Safe Centre is provided to deidentified microdata sets only for scientific purposes, which means that direct identifiers (e.g. unique identification numbers, phone number, e-mail address, etc.) are removed or replaced with artificial technical numbers. Numerous, ready-made standard datasets are available for researchers, e.g.: Labour Force Survey, Household Budget Survey, European Health Interview Survey, Time Use Survey, Population and Housing Census.

If no appropriate standard datasets are available for a given research purpose, specific datasets can be requested for a fee. For example in the field of live birth, death, marriage, divorce, fetal losses, internal/external trade, corporate tax reports, product-level trade data etc. See more details in the appendix.

5. IT solution

The following IT solution was applied in the new Safe Centre:

The researchers can interact with the central terminal (server) dedicated to the Safe Centre via clients. The central terminal and the clients operate in a virtual environment, the communication between them takes place over a public internet network.

All the data (input data sets, intermediate result, outputs) are stored on the central terminal. The researchers are able to open and read the data sets from the central server, write and run their program scripts with a statistical software (e.g. STATA, R, MS Office), but downloading or exporting any data or outputs from the server is not possible in this environment. The only form of external appearance of data is through the monitor of the computers.

If the researchers would like to bring a dataset or a pre-written program script into this environment, they must submit it to the staff of the Safe Centre, who carry-out the so-called input checking. If the staff approve the inputs, the files are copied to the appropriate destination on the central server.

Regarding the results created by the researchers in the Safe Centre, the colleagues of the HCSO fully examine the research outputs produced from statistical confidentiality point of view, prior to their release to the researcher. The HCSO provides access to the researcher only for those research outputs that have been checked and approved against statistical confidentiality.

Two VPN connections were configured: one between the servers of the HCSO (where the HCSO’s microdata sets are stored) and the remote access point of the HAcS Safe Centre, and another which establishes a connection between the HCSO’s monitors and the HAcS Safe Centre’s CCTV cameras.

The clients of the remote access point were given a permission to use the first VPN connection to log in and to operate on the Safe Centre’s environment. This channel is only open during the opening hours (9 am - 4 pm), otherwise the connection is terminated.

---

4 In the Safe Centre, so-called zero clients – which are the minimization of the thin client – are used. The parts of the thin client are reduced to get the zero client, which contains only an operation system with minimal capabilities, and has no hard drive nor CD/DVD player. The zero client completely relies on the resources of the server, thus it is just a simple intermediate device between the server and the user.
The server of the remote access point and the clients are based on the VMware Horizon View virtual technology. The access is secured with Firewall solution.

The CCTV server is available all day long (0-24). In this channel, only the access to the cameras is available.

In the above described system (which is depicted on Figure 3), only the 24 hours CCTV monitoring can grant the full security.

Currently 12 zero clients and 20 virtual computers operate in the new Safe Centre with statistical software STATA, R and MS Office.

---

5 See more details here: https://www.vmware.com/uk.html
6. The support activity of the HAcS regarding the Safe Centre

In this section, the role of the HAcS in the data access process and the support activities is pointed out, tasks that the colleagues of the HAcS carry out in the new Safe Centre.

In order to access the Safe Centre, the researchers need to make an appointment in the registration system created by the HAcS for this specific purpose. This system is operated and maintained by the HAcS, who give regular reports to the HCSO on the use of the Safe Centre by the researchers.

Furthermore, the HAcS keeps in touch with the HCSO continuously, tracks the microdata needs, does the administration and coordinates the process. They facilitate the tasks of the researchers regarding the Safe Centre, basically they act as a mediator between the researchers and the HCSO.

The HAcS played a vital role in the introduction of the new Safe Centre to the researchers, and they have been doing this task continuously since then. They monitor the CCTV cameras during the opening hours to avoid any incidents. If, for any reason an incident happens, they immediately report this to the HCSO and prepare a formal report. This report is forwarded to the HCSO for further processing.

However, the HCSO also has the opportunity and obligation to monitor the CCTV cameras and check in with the colleagues at place if they spot anything out of the ordinary.

7. Benefits of the new Safe Centre

There are many advantages of a Safe Centre located outside the NSI for both parties. We summarize these benefits in the following table and then describe 2-2 advantages in more depth.

<table>
<thead>
<tr>
<th></th>
<th>HCSO</th>
<th>Researchers</th>
</tr>
</thead>
<tbody>
<tr>
<td>The relationship between the two institutions is strengthened, the cooperation becomes more efficient</td>
<td>With cooperation and daily communication, the HCSO is notified much earlier about the emerging research needs and issues</td>
<td></td>
</tr>
<tr>
<td>As there is a constant contact with the users (researchers), their needs and problems with the datasets come to light earlier</td>
<td>The HCSO is able to get more experience, which enables to support the work of the Official Statistical Service more efficiently regarding the data requests for scientific purposes</td>
<td></td>
</tr>
<tr>
<td>The HCSO is able to get more experience, which enables to support the work of the Official Statistical Service more efficiently regarding the data requests for scientific purposes</td>
<td>Time-saving! The researcher can save the time of traveling to the HCSO’s Safe Centre. There was a 400% increase in Safe Centre usage when it moved to the HAcS building.</td>
<td></td>
</tr>
<tr>
<td>Wider use of the HCSO’s data</td>
<td>The research facility is able to provide the necessary microdata to its researcher community</td>
<td></td>
</tr>
</tbody>
</table>
Methodological benefits in the field of output checking: the only way to get experiences and to develop new methods is to check more and more outputs

“Output saving”: the research leader can check the outputs locally, and only the necessary outputs are taken out

| 1. Table - Benefits of the new Safe Centre from both side’s viewpoint |

Why is it worth to establish a Safe Centre in the premises of a research institution? First of all, it ensures a more efficient research for the researchers as they can continuously communicate and support each other. In case of the Safe Centre located in the HCSO, if an issue emerged, the researcher had to face it himself. Without the usual communication channels (internet, e-mail, phone), the problem could not be solved on site because no assistance (methodological, programming) from a more experienced researcher/supervisor was available. In the new Safe Centre, all the necessary help is in the vicinity, so no delays will occur in the research. This increase the effectiveness and also saves a lot of time, since the researcher does not need to travel remarkable distances in the city.

The new Safe Centre is output-saving as well. Beforehand, the leading researcher wrote the program scripts in advance, sent the assistant to run the codes in the HCSO’s Safe Centre and they took all the outputs created back to the research institute. In the new Safe Centre as it is located in the research institution’s building, all the researchers involved in the given research can immediately examine the outputs, and only the necessary ones are requested. This also saves time for the HCSO as less output checking is required, which at the end also saves costs on both sides.

From the HCSO side, the NSI takes into consideration and support the needs of the research community, which strengthen the relationship between the HCSO and its users. Its microdata sets are used in a wider range and not just the recent ones, but data dating back even to the 1970s are being researched.

There are benefits regarding output checking: the HCSO gets more and more experience and is able to continuously develop its methodology.

8. Costs

Setting up the Safe Centre costs about 450,000 € on the part of the HAcS. This amount covered the price of two high-capacity servers (owned by the HAcS but operated in the HCSO computing centre), establishment of the remote control system, and costs of the first set of data made available for research. This amount excludes the resources allocated for this purpose by the HCSO, and the respective workload of the Data Bank. The Academy pays an annual fee of about 30,000 € for server operation, preparation of the data sets, the cost of adding new waves to the existing data, and output checking. The costs have so far been covered by the central budget of the Academy of Sciences. The services are provided free of charge.
9. Current operation in numbers

In the following section, we describe the Safe Centre’s operation in numbers. It is important to mention the utilization of the Safe Centre. The best measurement for this is the number of reservations at the Safe Centre. We registered 6342 sessions in the Safe Center since 2014, the opening of the first room in the HCSO’s building. A total of 93 researchers working on 43 projects used the data since then. Moving the room to the premises of the HAcS resulted in a major growth of usage (3815 sessions between February 2018 and June 2019).

Considering the number of outputs in the last year, 105 were created. This number has changed to 51 (till the end of May, 2019).

![Number of outputs](image.png)

4. Figure - Number of outputs

10. Satisfaction

According to a recent survey conducted among users, the overwhelming majority of them is satisfied with the system, except the opening hours (9.00-16.00, 9.00-13.00 on Fridays) that most of them find too short. The HCSO sets the hours of operation so as to ensure that competent decision-makers be available in case direct intervention is required.

11. Experiences, recommendations and plans for the future

In order to make the new, joint Safe Centre more efficient, a common trust is needed from both sides. It is important that the research community communicate and stress the need for microdata sets, and that the HCSO should be very open to fulfill these needs.
The number of researchers and research studies should be numerous and the research activity ought to be continuous at the research institution to exploit sufficiently the possibilities of the Safe Centre. The framework of the Safe Centre and the collaboration between the two parties need to be flexible enough, so as to be extendable not just from information technology side, but from research institution side, i.e. other research institution should have the chance to use the new Safe Centre. This helps to avoid the situation of monopoly.

It is important to emphasize that the new Safe Centre helps to create a new, young generation of researchers where the data usage gets more and more relevant than before. The research institution gives lectures to the researchers where they can study statistical software used in the Safe Centre to analyze the data (e.g. STATA). These lectures are open to the colleagues responsible for output checking at the HCSO as well. This makes the process of output checking more effective.

As for the future plans, it is vital to identify the non-practical and rigid elements of the data access at the new Safe Centre. To do that, continuous monitoring and revision are required. This allows to increase the quality of service the Safe Centre can offer.

Part of our short term plans is to establish a new Safe Centre outside the capital city. Pécs is a very prominent candidate as this is the centre of the Institute for Regional Studies. If a Safe Centre is established in Pécs, all co sciences of HAcS get access to the research room.

12. Conclusion

The joint Safe Centre facilitates the work of the researchers, it enables them to make their research on very detailed microdata sets. It also guarantees benefits for the NSI, besides increasing its reputation, its data are used in a much wider range of users.

The framework of the new Safe Centre is based on the existing Safe Centre’s working model in the HCSO, complemented with a couple new standard processes. The most important part of the collaboration is that the HAcS has become part of the data access process, they form a bridge between the researchers and the statistical office. Their activity makes both sides life much easier.

It is essential to stress that although written common rules and agreements were made and are in force, such rules and agreements cannot be made exhaustive enough. Reality is always far more complex and cannot be anticipated. Hence, trust between the cooperating institutions is at utmost importance. We have to trust that each of us does what were agreed on and we also have to trust that we will find common, acceptable solutions on the problems we cannot foreseen yet.
Appendix
Data available for researchers in the new Safe Centre

— Microcensus (2005; 2016)
— 10% sample of the Population and Housing Census (2001; 2011)
— European Health Interview Survey (EHIS) microdata sets (2009, 2014)
— Time Use Survey microdata sets (2010)
— Divorce datasets (1977-2017)
— Mobility (1975-2017)
— Housing statistics (1999; 2003; 2015)
— Structural business statistics datasets (2006-2016)
— General practitioner’s activities (2001-2017)

References


Grim, R., Heus, P., Mulcahy, T. and Ryssevik J. (2009), Secure Remote Access system for an upgraded CESSDA RI.

Hungarian Central Statistical Office Strategy 2020; (2014)

Thaulow I. and Nielsen C., (2015), New Nordic model for researchers joint access to data from the Nordic Statistical Institution, Helsinki.