Improving Access to Official Statistical Data – Experiences of the Hungarian Central Statistical Office

Mr. Zoltán Vereczkei*
Hungarian Central Statistical Office, Methodology Department, Budapest, Hungary – zoltan.vereczkei@ksh.hu

Ms. Eszter Nagy
Hungarian Central Statistical Office, Legal Affairs and Administration Department, Budapest, Hungary – eszter.nagy@ksh.hu

Abstract

The Hungarian Central Statistical Office (HCSO) – as other National Statistical Institutes – is facing a large and steadily growing need for access to official statistical microdata. Hungarian and European researchers have a particular interest in modern, fast and transparent ways of access to official statistics. In order to meet present and future user needs for data access, the HCSO needed to adapt its data access tools as well as procedures. Satisfying user needs is of particular importance for the HCSO, therefore an internal development program has been launched in 2012 in order to evaluate and modernise the ways of data access and the supporting procedures. This modernisation relies heavily on European and international recommendations, good practices and the own experiences of the Hungarian Central Statistical Office. As a key part of the program, the data access channels of the HCSO have been redefined and improved and several internal IT developments have begun. As outcomes of this modernisation program, the HCSO has standardised the procedures for data access, systematised the data access channels, renewed the system of internal data access and data protection regulations with built-in methodological safeguards, restructured the internal organisation for data protection, introduced researcher accreditation, developed an internal IT tool to support management of data requests and increased the amount of publicly available information on data access channels and procedures. In order to further modernise data access and increase the quality of data access services, several additional internal developments are still being carried out in the fields of methodology, infrastructure, organisation, IT security and process management. The paper summarises the main drivers for the modernisation activities in the field of data access, the results of the finished and ongoing internal developments and the main lessons learned by the Hungarian Central Statistical Office.

Keywords: data access; microdata access; modernisation.

1. Introduction

Since its establishment in 1867, it is part of the mission of the Hungarian Central Statistical Office (HCSO) to apply data protection in the whole process of the statistical business processes. Parallel to the data protection considerations, providing access to official statistical data is one of the driving forces of compiling official statistics. In order to bring the two approaches into a balance, access to official statistics is provided with full compliance with data protection rules.

The basic principles for this balance remain stable and unchanged over the years, the user needs, however, are constantly changing and new advances in technology appear, generating new data access solutions and opening new horizons for access to official statistics. The HCSO, as National Statistical Institutes in general, faces a growing need from researchers for microdata access in safe environment and for new solutions to make data access more convenient and faster for them. In order to meet these new needs and to keep up with technical development, HCSO launched a set of modernisation projects on the domains of data access and data protection in 2012.

Apart from the internal motivations for changes and modernisation (both from NSI and researcher sides), international initiatives such as the Data without Boundaries project, the OECD Expert Group for International Collaboration on Microdata Access or the revision or EU Regulation No. 831/2002 had big impacts on the modernisation of data access solutions to the national official statistics of Hungary.

As a significant result of this modernisation, the data access channels to national official statistics have been systematised and reorganised and the internal IT-based operating environment has also undergone big changes.

2. Modernisation of data access channels and establishment of researcher accreditation

As a first step of the modernisation activity, the internal regulatory environment on data protection has been renewed, a confidentiality policy and a new internal regulation on data access have been developed, based on national and international needs and lessons learned. Part of this activity, the HCSO reorganised its data access portfolio, now consisting of six data access channels: access to tabular data, release of anonymised microdata sets, access to public use files (PUF), access in Safe Centre, remote access and remote execution. The last three channels are generally referred to as access in safe environment.

Access to tabular data and access to PUF data access channels are available to all users without any limitation: PUF are freely downloadable from the NSI website (with the acceptance ot terms of use) and tabular data can be requested by any user thus the scientific nature of the data request is not examined here.

The remaining four data access channels (access in safe environment and release of anonymised microdata sets) are available for researchers only for the purpose of conducting scientific research. Thus, in these cases, the data request is examined to ensure that the intended use is scientific research. It is an important characteristic of the Hungarian data access system that users with and without institutional background can have access to official statistical data stored in safe environment. However, researchers without institutional background may not access anonymised microdata sets. The motivation for keeping disclosure risk within manageable limits is the main driver for this restriction.

In order to ensure compliance with these requirements in practice, a system of researcher accreditation has been introduced in the HCSO as a new integral element of data access procedures. The aim of the researcher accreditation is to ensure that data requests are of scientific nature thus access to the requested official statistical data can be provided – given that the requested data is available – in line with the dedication of the NSI to support scientific research.

Supporting instruments of new standard data request templates, contract templates, confidentiality declarations, terms of use documents have also been developed and the NSI website content has been substantially extended with information on all data access channels and the core data protection documents both in Hungarian and English (e.g. the data confidentiality policy, the internal HCSO regulation on data protection and the HCSO IT security policy).

Figure 1. shows the framework of any data access process, supported by these instruments.

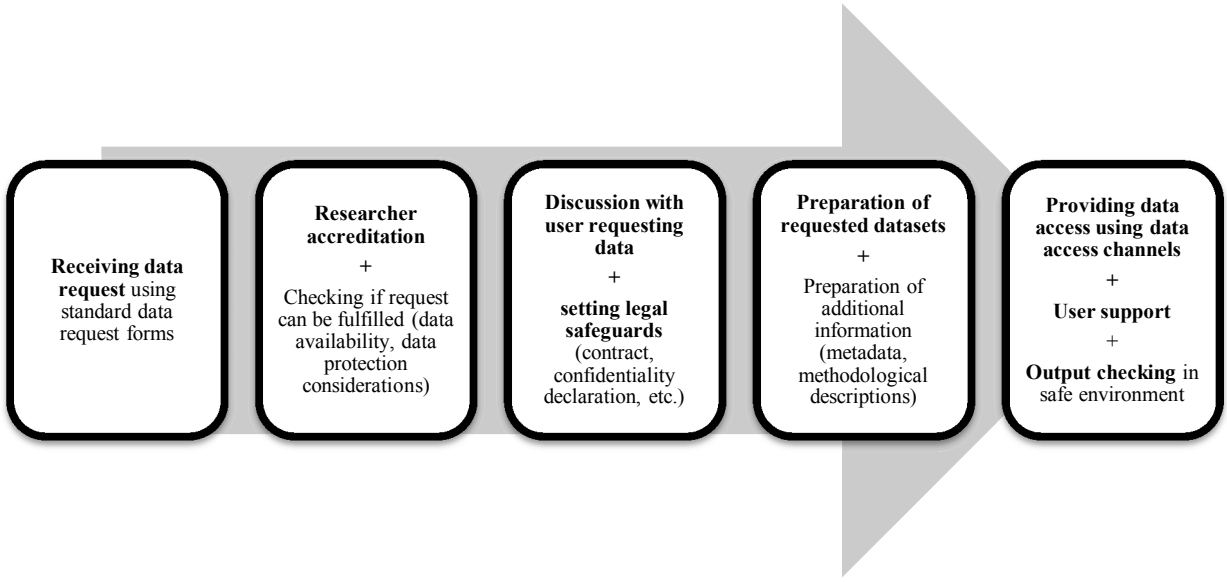

Figure 1. – Schematic model of data access process

To ensure that data access is provided in a service-oriented manner, the internal HCSO procedures have also been redesigned. As a response to the growing need for access in safe environment, the importance and the role of output checking activities in the statistical business processes of HCSO is constantly growing. Started with the reorganisation and modernisation of the HCSO Safe Centre a few years ago, the output checking activity has been integrated into the HCSO processes during the last few years. Development of an internal guideline on output checking (focusing both on methodological and process-oriented elements) has begun and several advancements to the data access procedures related to safe enviroment have been made. This activity resulted in the creation of a publicly available guide for researchers (focusing not only on output checking but also on the main aspects of data access and data protection in general) and updated rules of procedures for the Safe Centre.

Altogether with the new provisions on output checking, the new internal data protection regulations – in line with the HCSO Quality Guidelines – define the mandatory provisions on protection of tabular data (both primary and secondary cell suppression), the general approach on anonymisation of microdata sets (definition of disclosure scenarios and key variables with the estimation of disclosure risk) and the output checking activity.

3. Integration of recent developments into data access procedures

For the development of the Hungarian researcher accreditation system, information collected by the Data without Boundaries project and the current Eurostat pratice – based on Regulation No. 557/2013. – were closely examined. The Hungarian researcher accreditation system is now a new and important element of the data access procedures of the HCSO, in line with the current international practices and recommendations.

One of the key areas of the recent (and partly still ongoing) developments is the redesign of access procedures in safe environment. The Safe Centre of the HCSO has been fundamentally rethought and modernised, resulting in exponentially growing data requests by researchers for this mode of access. The current trends still show an ever increasing need for this data access channel, resulting in the constant need for further development and expansion of the safe environment. In order to meet these current and exptected future needs, the HCSO started building its own remote access system. As part of this modernisation line, a concept of the Hungarian remote access system has been developed under

project MIRA, building strongly on the results of the ESSnet[1] DARA (Decentralised and Remote Access to Confidential Data in the European Statistical System (ESS)). In the framework of the development of the Hungarian remote access system to official statistics, a pilot remote access point in Szeged has already been established and new remote access points are under construction.

As a response for growing needs for microdata files for teaching purposes and in order to further support the preparatory work of researchers in the safe environment, the HCSO has started developing PUFs, based on European and international recommendations and guidelines. Both the finished and the ongoing methodological developments on PUFs result in so-called test PUFs and teaching PUFs, available for download on the HCSO website.

In order to support data access procedures and to further systematise data protection approaches and the applied practices office-wide in the HCSO, the Data Protection Board has been established. The Board serves as a presidential advisory body and also provides practical recommendations to any employee of the HCSO in issues related to data protection, data access or both. The Board consists of the representatives of the legal, methodological, IT, dissemination and IT security areas and chared by the data protection commissioner of the HCSO.

4. Development of internal and external IT tools supporting data access

In order to support the transparency and documentedness of the data access procedures, a new internal IT tool, the Internal Data Request Management System (ADKI) has been developed. With the ADKI IT tool, the whole data access procedures are tracable – as also shown on *Figure 2.* – and all related documents are available for HCSO employees (data requests, e-mail exchanges, price bids, contracts, confidentiality declarations, methodological descriptions). One of the main benefits of the ADKI is that based on the rich information available in the system, the management of disclosure risk of a given data request can be more objectively estimated.

The currently ongoing or recently finished additional IT developments will more systematically integrate ADKI into the IT architectural environment of the HCSO. The newly developed Contact Center system will enhance the communication activities with the users and provide e-form facilities to send data requests, further easing the use of the service (the current system is based on editable PDF files sent via e-mail to the HCSO information service). With another newly developed intergrated HCSO IT tool, called KARÁT (Integrated Data Transmission System for Secondary Data), all transmissions that contain data sent to users based on data requests or agreements will be managed by this system, further enhancing the level of IT security, ceasing all data transmissions by direct e-mails and providing a secure channel for sending the data.

[1] An ESSnet is a project organisation established within the European Statistical System, based on cooperation of several NSIs in order to establish results for the whole ESS, based on common interest.

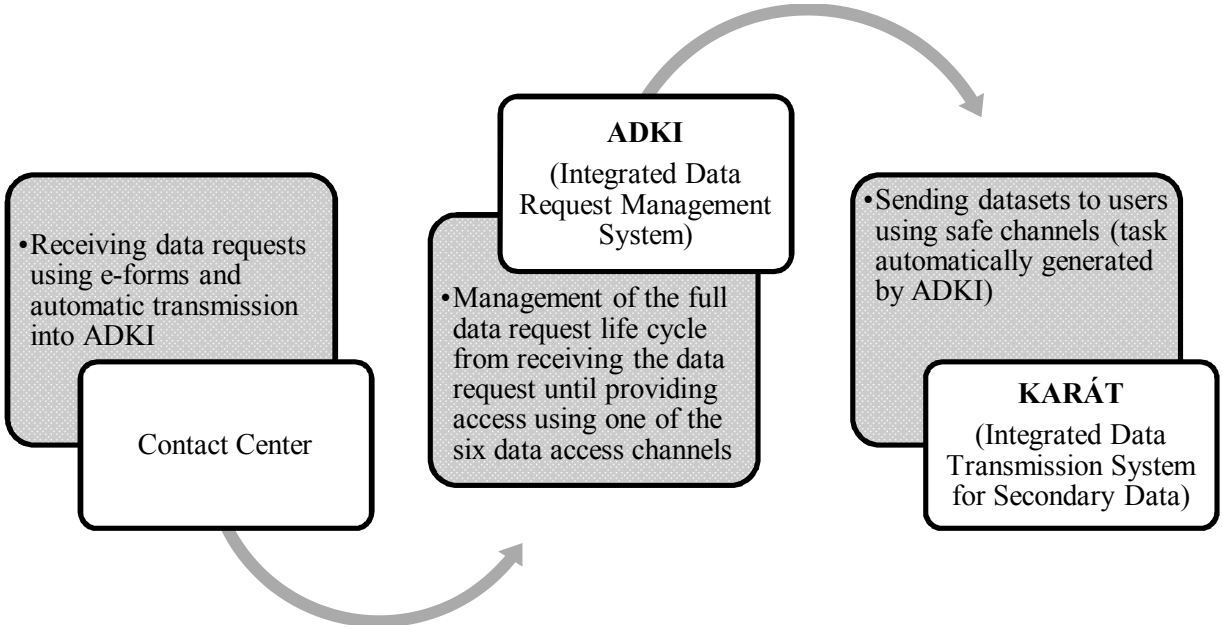

Figure 2. – Connections of intergrated IT tools from the perspective of data requests

As a result of these IT tool developments, the ADKI is now an intermediate system between the Contact Center and the KARÁT, receiving data request information automatically from the e-forms via the Contact Center and managing the transmission of the prepared dataset to users (based on an automatism) via KARÁT.

5. Key lessons learned from recent developments

The modernisation of data protection and data access solutions in the HCSO has been an important development line for the HCSO in the last 4 years and is still one of the most rapidly developing areas. The already established deliverables of the modernisation activity in the fields of data access and data confidentiality have both improved the access to official statistical data and established the necessary supporting instruments to further enhance the concerned internal HCSO processes.

With the expansion of data access facilities (both with the continuous modernisation of the Safe Centre and the opening of new remote access points) the number of data request is constantly growing. The strenghtening role of access in safe environment is obvious, making output checking activites more and more core tasks in the statistical business processes. With the more prominent role of output checking, the statistical disclosure control portfolio is expanding: new SDC methods appear while the 'traditional' SDC methods applied for tabular data and microdata anonymisation are still key areas for the methodology of data protection.

6. The way forward

In order to prepare for the challenges of the upcoming years in the fields of data access and data protection, the Data Confidentiality Board has prepared an action plan for HCSO on data protection that defines the main areas for development in the years to come until 2020.

Based on key lessons learned, the HCSO is committed to open new remote access points with the further development of the remote access system and the Safe Centre. To support the associated output checking activity and enhance methodological guidance, the HCSO will produce its own output checking guideline in 2015. Connected to the further development of data access in safe environment, the HCSO will prepare a business case for the redesign and further development of its remote execution facilities.

Apart from the further development of the data access channels, the key areas for the future developments – in line with the recommendations of the OECD Expert Group for International Collaboration on Microdata Access – are the investigation of the possible integration of metadata standards (with special attention given to DDI) to improve information available for users on microdata sets, the further expansion and standardisation of data access procedures with the involvement of the members of the Hungarian National Statistical Service and further enhancing cooperation with the researcher fora.

The recent and still ongoing European and other international developments in the fields of data access and data protection could significantly boost the need for the modernisation of data access solutions. Providing detailed information on data access procedures in English is the basis for information exchange with other NSIs in the ESS and beyond and is a prerequisite for the harmonisation of the practices of different countries and international organisations.

6. Conclusions

The Hungarian Central Statistical Office launched a modernisation line in the domains of data access and data protection in 2012 with the goal to redesign and further enhance data access facilities in full compliance with data protection considerations. As results of this activity, several improvements in the data access procedures have already been implemented, including the systematisation of data access channels, development and standardisation of their supporting instruments and the establishment of a system of researcher accreditation.

Apart from these large scale developments, the internal HCSO procedures on data access and data protection have also been redesigned to better support the changing environment. Special attention is to be given to the constantly growing needs for access in safe environment with the output checking activity becoming more and more prominent is core statistical business processes.

The results presented in this paper show the outcomes of the development produced so far. As data access is a key area, especially for the user group of researchers, the further development of data access procedures will follow in the upcoming years.

References

ESSnet DARA deliverables (2013). http://www.cros-portal.eu/content/decentralised-and-remote-access-confidential-data-ess-dara

Data without Boundaries deliverables. http://www.dwbproject.org/

OECD Expert Group for International Collaboration on Microdata Access – final report (2014). www.oecd.org/std/microdata-access-final-report-OECD-2014.pdf