

Proposing the Model for Croatian Remote Access Safe Centre for Statistical Microdata

Martina Poljičak
Croatian Bureau of Statistics
Ilica 3, Zagreb, Croatia
poljicakm@gmail.com

Hrvoje Stančić
Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
hstancic@ffzg.hr

Summary

The data collected in National Statistical Offices (NSOs) present a valuable source for the research community. These data, when put to some secondary analyses, bring new benefits for society. The growing need for researchers' easy access to publicly funded data is recognised in the European Statistical System (ESS). There are on-going efforts for establishment of pan-European network for Remote Access (RA) to Microdata in official statistics and there are several projects concerned with these issues coming along. In this research paper the authors will explore the current state of establishment of infrastructure for researchers' RA to statistical microdata in ESS and they will propose a model for development of infrastructure for researchers' RA to statistical microdata in Croatia with the aim of integrating the Croatian system for RA to statistical microdata in the currently developing ESS's infrastructure for the same purpose.

Key words: European Statistical System (ESS), Statistical Microdata, Remote Access (RA)

Introduction

Through history it has been a common practice for European countries to provide public only with aggregated statistical information. These aggregated data, also known as macrodata, present results from publicly funded statistical surveys. Data from which macrodata are produced, i.e. microdata, would reside in the premises, i.e. on the servers of organisations performing these surveys, or some partner archiving organisation responsible for safe preservation of statistical data. However, the prevailing trend and orientation of global politics is changing towards opening access to publicly funded data, collected in publicly

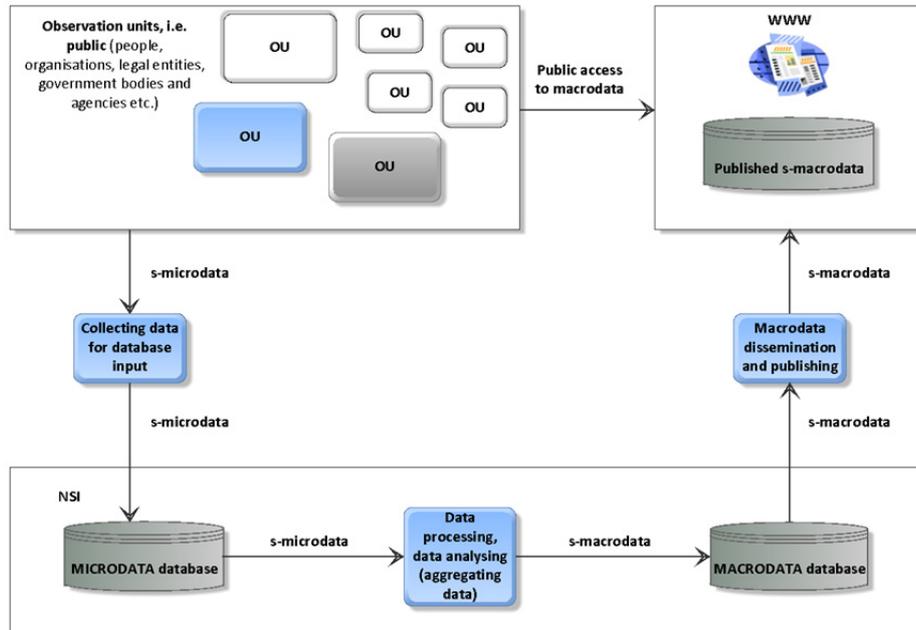
funded projects. Yet, another trend in information and communication technologies is towards turning closed organisations' information systems into open systems, available on the Internet through a set of web services, i.e. orientation towards service oriented architecture (SOA) and service applications available for global society users. The same trend is introduced in statistical information systems management. The main objective in development of information systems in statistics is set to harmonisation of statistical processes, standardisation of procedures and establishment of cooperation between different countries through means of information and communication technologies and standardised datasets. New systems, oriented towards opening access to raw data for the public, are on their way to become present in countries all over the world. These systems are considered necessary in order to have more transparent and effective society. This new approach will bring many changes and will have a big impact on everyday life in today's knowledge-driven society. Special concerns in statistics area are the legal compliance of the new social system and safe implementation of the needed services, especially due to the fact that different countries deploy different legislatives about statistical microdata confidentiality. Statistical microdata are an example of confidential datasets, so they will probably never be completely open for public use. However, they can be opened to some extent¹ for researchers' use.

Data throughout statistical surveys

To be able to provide statistical indicators about society and environment to the public, NSOs in countries collect microdata about statistical observation units (people, organisations, events, etc.). NSOs process those microdata about statistical observation units and produce aggregated data, i.e. statistical macrodata. Statistical macrodata are disseminated by various publication channels². Statistical macrodata are publicly owned goods. The system presented by the model in the Picture 1 should be useful for explaining micro and macrodata-flows between public and NSO. The authors have illustrated the basic model of NSO, collecting microdata from public and in return providing public access to statistical macrodata. Amongst collecting microdata through surveys from public, there are some other possible sources of microdata for NSOs, like administrative registers, providing microdata input. Having that in mind the model presented in the Picture 1 should be perceived as a basic model of collecting statistical microdata, presenting basic methods of collecting statistical microdata, i.e. statistical surveys.

¹ Usually anonymised statistical microdata are opened for use to the research community.

² Usually by printed or web editions or by interactive database warehouses available on the web which enable browsing and analysing statistical macrodata.



Picture 1. National Statistical Office providing public access to macrodata (basic model of NSO)

NSOs have legal right³ to collect confidential data about statistical survey units (persons, institutions, events, etc.) as they are objects (input data) for statistical surveys.

Microdata Confidentiality

Microdata are protected by legislatives which differ from country to country.⁴ As stated in Data Without Boundaries (2012a) “ (...) highly-detailed and confidential microdata (...) should be understood as data about individual statistical units that, due to the level of details provided, contain the risk of disclosure for the specific statistical unit. Therefore such data can only be accessed under technical, organizational and statistical restrictions.” Definitions of confidential

³ These rights can only be practiced as stated in legislatives covering statistical survey in question and only persons involved in the statistical survey have access to data in question. All staff members involved in data processing in NSOs and therefore having access to confidential data are legally obligated to secure and prevent data from being disclosed to other persons during and after the period of involvement in statistical processes, i.e. statistical surveys, and are especially forbidden to use statistical data for nonstatistical purposes of any kind.

⁴ Data classification should be and it already is in place in many countries.

microdata records are not limited just on data presenting one statistical unit but “definitions of confidential microdata depend on national legislation, the design of the survey (sample size, included variables) and the means of access (e.g. Remote Access, Safe Centre, Download).”⁵ This means that small sets of records can be considered too revealing and the possibility of disclosure of a single record in a recordset could be the reason for classifying the whole recordset as confidential.

Researchers’ access to Microdata

Research communities and governmental bodies are using statistical microdata as they are useful when put to some secondary use and made available to researchers who are in turn able to produce new and meaningful representations for the society. In *Data Without Boundaries* (2012a) a clear conclusion is made that “only real Remote Access (RA) systems would meet the cumulative challenge of the high demand for confidential microdata, limited resources inside the NSIs (...) and researcher’s needs.”

ESS projects for development of RA to Microdata

As presented by the results in *Data Without Boundaries* (2012a) some European countries, namely Denmark, France, Slovenia, the Netherlands, Sweden and United Kingdom, offer researchers RA service to their official statistical microdata. Also, there is a Safe Centre in Eurostat⁶ in Luxemburg where researchers can remotely access microdata from some statistical surveys of several EU member countries. At this moment these procedures are very complex and time consuming and there is a need for establishing a more cooperative, inter-connected and user-friendly system. Also, there are running projects, like ESSnet’s “Decentralised and Remote Access to Confidential Data in the ESS” - DARA, concerned with the establishment of European Safe Centre for Access to statistical microdata in all Europe’s NSOs and partner organisations, as well as with connecting Safe Centres in countries with the servers in Luxemburg. The project “Decentralised Access to EU microdata sets” ended in January 2010 and it produced a feasibility study for this establishment and was also a trigger for the DARA project. Eurostat is in the process of establishing “Secure Infrastructure for Confidential data access and sharing” (VIP-SICON) project to build a remote access system based on Citrix and the software development for this purpose has just begun, so at this moment it is still not evident how this system will operate. The *Data Without Boundaries* project is aimed at developing an easy and equal access rights and infrastructure to official microdata for researchers

⁵ *Data Without Boundaries* (2012a).

⁶ Eurostat is the Statistical Office of the European Union, i.e. central statistical authority in the ESS.

for the European Research Area through coordination of existing infrastructures, Council of European Social Science Data Archives (CESSDA⁷) and the ESS.

Available RA Centres in Europe

A definition of Remote Access to microdata, as presented in Data Without Boundaries (2012a), is „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 the real, highly detailed and confidential microdata.“ The first country to offer their researchers RA to microdata was Denmark in 2002, followed by UK and Slovenia in 2004 and Sweden in 2005. In 2006 the Netherlands joined this community and two more countries in 2009, namely France and again UK⁸ introduced microdata RA, followed by Germany in 2011. Other countries still have to develop similar systems based on existing models with the aim of being compatible to others, as the movement towards cooperation in statistical area is evident and expected by authorities like EU Commission⁹ and Eurostat.

Common and different RA areas in available centres

Although, at this time, there is no case of two exactly the same RA centres in Europe – they all, according to the study performed and results presented in Data Without Boundaries (2012a), have some common areas in the way they are assembled and organised to work. Firstly, in all RA centres confidential data does not leave the research data centre. Users basically connect to RDC's¹⁰ server with thin clients¹¹ or via regular PCs, which are only able to trigger execution of commands on the protected RDC's server, where data is processed. Secondly, after being processed on RDC's server, data are sent – not directly to the user's PC but – to an RDC location. This location is usually user's outbox folder in RDC where the data stays until examined by an NSO's staff member. The data can be sent to a researcher only if approved to be safe for sending. All data transmitted to users are kept for some period in the archive of the RDC (3 months or so). At all the RDC's there is a process of accreditation before a re-

⁷ CESSDA at the moment has 20 member countries in Europe.

⁸ Different service then before in 2004.

⁹ EU Commission, along with Eurostat, has a big impact on the development of official statistical systems in Europe. The two organisations are active in introducing standards and establishing harmonised statistical systems in Europe.

¹⁰ RDC – Remote Data Centre.

¹¹ “A thin client is desktop computer that is only used to provide input and output to the user. The calculations are done on a server. According to that, thin clients are equipped with a limited operating system that allows only restricted use of functionalities (Data Without Boundaries (2012a)).”

searcher can get access. For this purpose passwords are used, given IP-ranges, smart cards with fingerprints, i.e. biometric technologies¹², smart cards that create one time password (OTP), hardware components¹³ or token with RSA SecureID. Most of the RDCs use more than just one of these methods for accreditation. All RDCs have sophisticated user account management and use Microsoft active directory for user management. The internet (or private government network¹⁴) enables remote connection between researchers and data, with relatively standardised technologies for encryption of communication. Some of the applied encryption methods are SSTP, SSI-VPN, Cisco-tunnel, RDP (Microsoft) and ICA Basic encryption (Citrix). Surveillance is carried out during research activities (methods like monitoring sessions, logs, log-on and log-off information etc.). At RDC's standard statistical software is usually made available free of charge (SAS, STATA, SPSS). Additional software is possible, but users have to cover the cost.

Building an integrated RA Centre for Microdata in Europe

Combined statistical microdata from various Data Centres available across Europe using RA services could be a precious source for researchers conducting comparisons and drawing important conclusions throughout the Europe, thus enabling better informed political decisions for society and economy. All this potential should not be neglected, especially in today's knowledge-driven economy. Many people agree on the notion that the so called "online borders" between European countries should be removed, thus enabling cooperation and even global proactive involvement of society, using global microdata sets from different countries. It is obvious, however, that these changes should be introduced with careful and reasonable consideration of possible benefits and possible threats for the society in the long term, and especially careful planning of new legal framework. Yet, at this moment researchers are facing complex procedures and have to be prepared to invest a lot of time and effort when in need of statistical microdata as they are available at the safe centre in Eurostat, Luxemburg with the exception of countries enabling this kind of remote access locally¹⁵ to their citizens, and yet another different set of rules is in place for the "noncitizens". The process of getting the authorisation privileges is rather complex and it takes some time before users can use the data. European Commis-

¹² Biometrics is being used currently in France.

¹³ In France the so called SD-box-hardware certificate device.

¹⁴ Private government network for researchers' RA to statistical microdata is used in UK only at this moment.

¹⁵ These countries are discussed in this paper, but they only provide access for their citizens, with an exception of Germany, which also provides remote access from US to the data in Germany, for a few US citizens involved in projects at IAB in Germany.

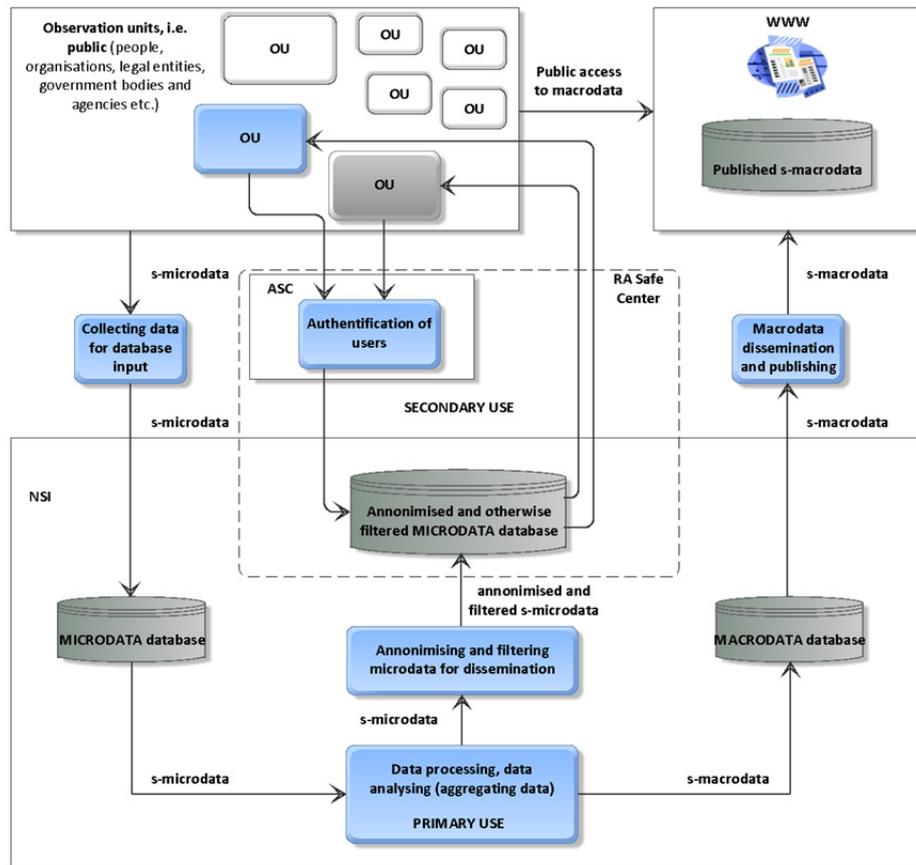
sion is on its way to simplify all these processes by implementing international e-signatures and international e-certificates on the Internet. It is presumable to expect their integration into governmental services for public over the Internet. There are some projects coming along at this moment with the aim of changing the state of RA to microdata in Europe. ESSnet projects deal with the problem of establishing an RA to microdata in Eurostat Safe Centre from local NSOs in European countries. DARA will deliver a running system connecting Safe Centres from countries in Europe to microdata at the Eurostat in Luxemburg. It already had two successful connections established from two countries in Europe until now. The project is planned to end in October 2013.

Proposal of the model for microdata RA in Croatia

The main statistical body of Croatia is Central Bureau of Statistics (CBS). Since Croatia became EU member in the July 2013 Croatia also took new obligations to fulfil requirements of the EU, Eurostat and other organisations involved in the ESS establishment. At the moment of writing this paper, Croatia is not a member of the CESSDA association. Following the global trend toward open government initiatives Croatian government has signed Open Government Partnership agreement in August 2011. Open Government Partnership, as stated in their web release, is a new multilateral initiative that aims to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance.¹⁶ Recently signed G8 Open Data Charter¹⁷ states that “access to data allows individuals and organisations to develop new insights and innovations that can improve the lives of others and help to improve the flow of information within and between countries. (...) Today, people expect to be able to access information and services electronically when and how they want. Increasingly, this is true of government data as well. (...) Open data can increase transparency about what government and business are doing.” Further in the Charter the G8 countries state that they will “establish an expectation that all government data be published openly by default (...) while recognising that there are legitimate reasons why some data cannot be released; (...) release high-quality open data that are timely, comprehensive, and accurate. To the extent possible, data will be in their original, unmodified form and at the finest level of granularity available; (...) make sure that data are fully described, so that consumers have sufficient information to understand their strengths, weaknesses, analytical limitations, and security requirements, as well as how to process the data; and release data as early as possible.” Also, they recognise “statistics (National Statistics, Census, infrastructure, wealth, skills)” as “areas of high value, both for improving our democracies and encouraging innovative re-use of data.”

¹⁶ Open Government Partnership, August 2011, <http://www.opengovpartnership.org/about>.

¹⁷ G8 Open Data Charter, UK Cabinet Office, 2013.



Picture 2. National Statistical Office providing access to macrodata for public users and anonymised and filtered microdata for researchers' RA and secondary analysis purposes

Remote Access to statistical microdata is one way to accomplish these goals. Croatia is an example of country without RA service centre for microdata in statistics. It is reasonable to expect that Croatia will be involved in developing Safe Centres for RA to microdata in the country connecting them with other Safe Centres in the ESS in the near future. Having that in mind, in the Picture 2 the authors are presenting a model for microdata RA Safe Centre for Croatia. In the proposed model shaded observation units present a researcher and a governmental institution employee seeking access to microdata. As illustrated, these users are able to get access to statistical microdata after passing through security checking by Authentication Service Centre (ASC). Considering the complex procedures for data classification in official statistics, which vary from

one statistical domain to the other, it is reasonable to expect a considerable effort will be invested in preparing automated, i.e. computerised, procedures for data classification in RA systems in order to have a fully automated system in place. For RA to microdata special authentication services must be put in place. Croatia has already established a similar service, called AAI identity, at the Croatian Academic and Research Network (CARNet). AAI provides electronic identity authentication service for members of the research community. However, Open Data initiatives are aimed at opening publicly funded data to all public members, not just to the researchers and government officials. If those systems are to be introduced and made available through a set of online web services, it is reasonable to expect that in the end these systems will be very complex, especially in the security layers implemented and authentication services used for various user profiles. Certainly, they could benefit from a planned introduction of NIAS¹⁸ in Croatia and this system should be developed with the aim of being compatible with various systems, hopefully it will be useful for implementing in researchers' authentication for statistical Microdata RA. It would be especially wise to think beyond the borders of Croatia. In that sense Croatia, as a member of ESS, should join the combined network for establishment of RA to statistical microdata in the ESS.

Conclusion

The full potential of statistical data is not derived by using aggregated data, i.e. macrodata. It should be possible to move beyond and to be able to connect, at least research community members, to safely protected statistical microdata using RA services and enabling easy access. Using that data, which possibly hold many undiscovered secrets and relationships, could introduce lots of benefits for society and empower individuals and organisations. In the last ten years outstanding progress has emerged in some European countries, and different countries have developed and implemented different sets of network infrastructure, protocols and ways of researchers' RA to statistical microdata implementing various ways of authentication of users and providing security for accessible statistical microdata. It is reasonable to expect future involvement of Croatia in this area as Croatia has just become a member of the EU and a full member of ESS. We have proposed the model for development of Safe Centre for RA to statistical microdata in Croatia, critically stressed the importance of harmonising applied procedures and careful introduction and development of system for microdata RA in Croatia with the aim of integrating this system into the ESS's system for RA to statistical microdata.

¹⁸ NIAS – National Identification and Authentication System.

References

- Brandt, Maurice. Decentralised and Remote Access to Confidential Data in the ESS (ESSnet DARA), Overview and State of the Art. 2012. http://www.safe-centre.info/wp-content/uploads/2012/07/ESSnet_DARA_Overview_Maurice_Brandt_20120328.pdf (05/02/2013)
- Brandt, Maurice; Zwick, Markus. Improvement of data access – The long way to remote data access in Germany. 2011. http://www.forschungsdatenzentrum.de/publikationen/veroeffentlichungen/fdz_arbeitspapier-39.pdf (05/02/2013)
- Council of European Social Science Data Archives (CESSDA). 2013. <http://www.cessda.org/about/> (05/02/2013)
- Data Service infrastructure for the Social Sciences and Humanities. About Dasish. 2013. http://dasish.eu/about_dasish/ (05/02/2013)
- Data Without Boundaries. Deliverable D4.1. Report on the current state of the art of current SC in Europe. 2012a. http://www.dwbproject.org/export/sites/default/about/public_deliverables/d4_1_current_sc_in_europe_report_full.pdf (05/02/2013)
- Data Without Boundaries. Deliverable D5.1. Report on concept for and components of European Service Centre for official statistics, 2012b. http://www.dwbproject.org/export/sites/default/about/public_deliverables/d5_1_european_service_centre_report.pdf (05/02/2013)
- Desai, Tanvi; Ritchie, Felix. Effective Researcher Management. 2009. <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2009/wp.15.e.pdf> (05/02/2013)
- Desai, Tanvi. ONS Virtual Microdata Laboratory: Strategic Recommendations 2011-2015. 2011. http://www.esrc.ac.uk/search/search-page.aspx?q=%3A*¤t_page=1&items_per_page=10&sort_order=relevance&filters=on&author=Tanvi%20Desai&tab=outputs (05/02/2013)
- ESSnet Home of DARA. ESSnet on Decentralised and Remote Access to Confidential Data. 2013. <http://www.safe-centre.info/> (05/02/2013)
- European Commission. Commission Regulation (EU) No 520/2010. 2010. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:151:0014:0015:EN:PDF>
- European Commission. Commission Regulation (EU) No 557/2013. 2013. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:164:0016:0019:EN:PDF>
- European Commission. Digital Agenda for Europe, A Europe 2020 Initiative. 2010. <http://ec.europa.eu/digital-agenda/digital-agenda-europe> (05/02/2013)
- European Strategy Forum on Research Infrastructures (ESFRI). Strategy report on Research Infrastructures, Roadmap 2010. 2010. http://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf#view=fit&pagemode=none (05/02/2013)
- G8 Open Data Charter, UK Cabinet Office, 2013, <https://www.gov.uk/government/publications/open-data-charter>
- Horton, Laurence; Katsanidou, Alexia. Purposing your survey: archives as a market regulator, or how can archives connect supply and demand?. 2011. http://iassistdata.org/downloads/iqvol35_horton.pdf (05/02/2013)
- Manifesto The New Global Data Generation. Simplifying and Guaranteeing Access and Sharing in e-Science Scenarios. 2012. <http://dasish.eu/manifesto/Manifesto2012-03-14.pdf> (05/02/2013)
- Open Government Partnership. Commitment of Croatia. 2011. <http://www.opengovpartnership.org/countries/croatia> (05/02/2013)
- Ritchie, Felix. Secure Access to confidential microdata: four years of the Virtual Microdata Laboratory. 2008. <http://www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/no--5--may-2008/secure-access-to-confidential-microdata--four-years-of-the-virtual-microdata-laboratory.pdf> (05/02/2013)
- Vrana, Radovan. Supporting e-Science: Scientific Research Data Curation. 2011. <http://infoz.ffzg.hr/INFuture/2011/ConferenceProgram.aspx> (05/02/2013)