CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Confidentiality

1-3 December 2021, Poland

Microdata access services coping with COVID-19 lockdown.

Natalia Volkow (National Institute of Statistics and Geography)

natalia.volkow@inegi.org.mx

Abstract

The paper presents the changes we undertook to the service provided by the Microdata Laboratory during COVID19 lockdown in order, to be able to support our users. The personnel of the Laboratory start home office since the end of March 2020, by May we realized the end of the pandemic would not come soon. We invite user to use Remote Processing, but for those using Microdata Laboratory it was not an option, because they needed to use the data they have already worked. We develop a new service Remote Processing in User´s Virtual Desk, this allowed users to use the information of their virtual desk and carry on their analysis even being in lockdown to support academic research and the definition, operation, or evaluation of public policy.

Microdata Access Services Resilience in COVID19 Lockdown

Natalia Volkow*

* National Institute of Statistics and Geography, natalia.volkow@inegi.org.mx

Abstract: The paper presents the actions taken to be able to provide access to microdata for research purpose during the lockdown needed by the COVID19 pandemic. The local context has special characteristics that had to be taken in consideration to guarantee that microdata access solutions safeguard the legitimate use of statistical information, only for research purpose or to support the definition, operation or evaluation of public policy.

1 Background

National Institute of Statistics and Geography (INEGI) was created in 1983 by the first statistical law published in 1980, within this legal framework there was no explicit remark regarding microdata access. It was not prohibited, so the Institute provided access to microdata of household surveys and Household and Polulation Census deliver with cost through external memory devices. To microdata to economic statistical projects, the only way to have access, was that the user had a person he knew that worked in the Institute that will allow him to go inside the building and guide the personnel of INEGI of what processing he needed. It was an *ad hoc* arrangement based on personal relation, if you did not have the personal connection, you could not have access. At that time the Institute also offered with cost, the service of special tables generated upon the request for all type of users.

In 2008 the National of Statistical and Geographic Information System Law was published it provided technical autonomy to the Institute and explicitly mention microdata access. Article 100 establishes that the Institute, following best international practices, will provide access to microdata, recognizing the importance of providing access to statistical information at lowest level of disaggregation, safeguarding confidentiality. The statement that the law includes is broad, so it allowed the organization of the microdata access services with good room of manoeuvre. The new law also established that the Institute will be ruled by a Board of Governors, with five members, one of them is the president of the INEGI. During the first session that the board was set up, it authorized the terms and conditions in which the Institute will provide access to microdata of all the statistical projects carried out by INEGI.

The agreement of the Board of Governors established that the rules regulating microdata access will vary depending on the type of statistical project:

1. For household surveys, government censuses, statistics of social data of administrative registers, the samples of the household and population censuses the access to microdata will be provided direct from the Institute web site by free download of the files of the household surveys that contain raw microdata. No registration is required. The confidentiality is kept by limiting to municipality (third geographical level) the location of each unit of observation from the random sample. In certain surveys location is provided at state (second geographical level). If the user needs the location of the microdata at lower geographical level, then it must be through the Microdata Laboratory.

2. The access to microdata from economic surveys and censuses (of establishments and agricultural units) and to the population and household census will be provided for free through Microdata Laboratory and Remote Processing only for research purpose or to public servants that are defining, operating, or evaluating public policies. The Institute also kept in operation the regular public service of processing specific tabulates, upon request with charge.

For the development of the microdata Laboratory INEGI we studied best international practices. A relevant gain from revising what other countries and the European Union were doing, was that we include in our microdata access regulation, since the beginning, transborder access.

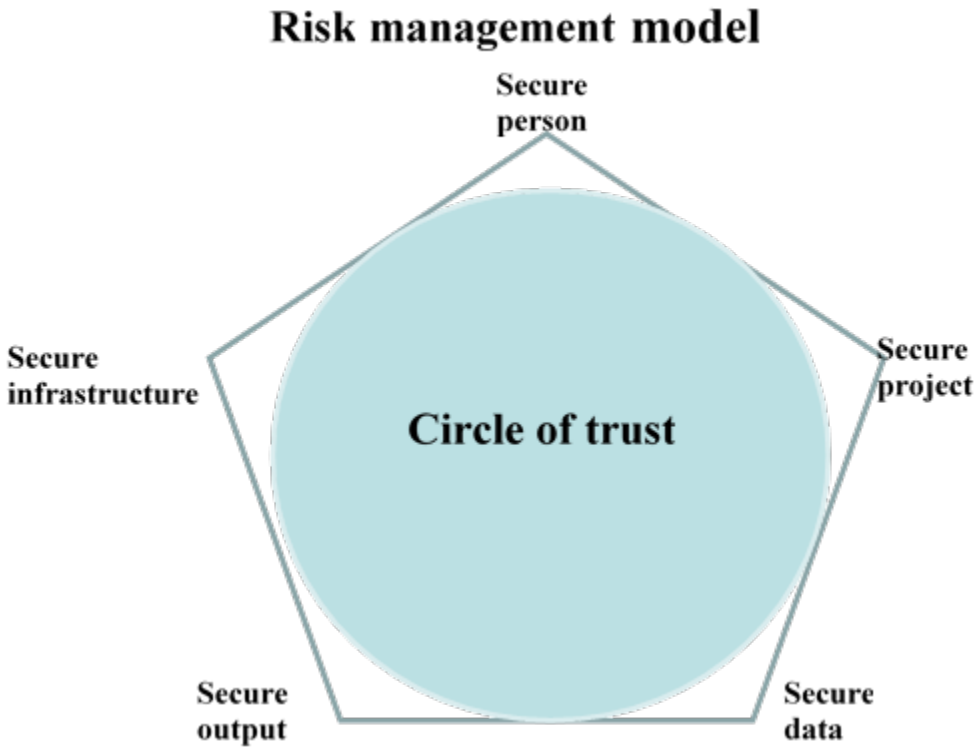
Origin of research project by country of institution of user affiliation	**Research projects**
Mexico	1412
United States of America	273
United Kingdom	52
Spain	27
France	25
Germany	23
Canada	20
Denmark, Netherlands & Italy	7
Chile, Colombia,	6
Sweden	4
South Korea & Japan	3

Argentina, Australia, Hong Kong, Ireland, Peru, Switzerland & Venezuela	2
Austria, Belgium; Brazil, China, Costa Rica, Finland, Ecuador, Malasia, Norway& Uruguay	1

From August 2011 to September 2021

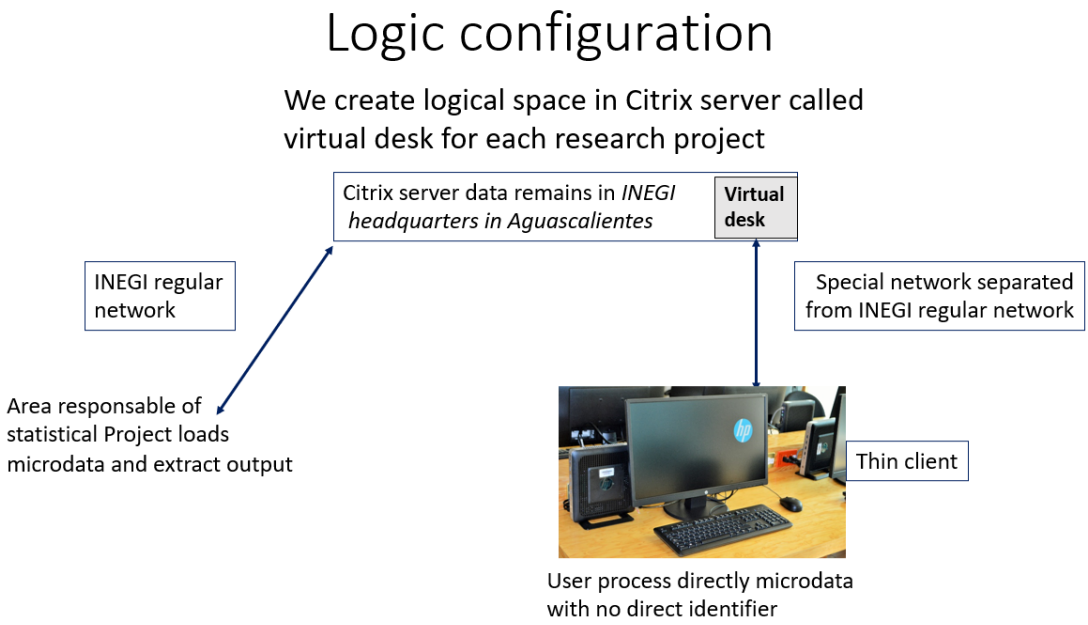
2 Microdata Laboratory

For the definition of the Microdata Laboratory service we follow the 5 ”S” model of UK Office of National Statistics, that is considered best international practice and was feasible to implement considering INEGI ICT infrastructure and local conditions.



INEGI developed the technical infrastructure of the Laboratory with the same type of technological platform that the Secure Data Service of the UK operates in its Secure Data Service, with a Citrix system. The platform allows the creation of a virtual space in which only authorized users with specific privileges can enter the secure enclave through terminals that are connected to a Citrix system. The personnel of the areas that generate information can access to

input the data need for a research project of a specific user or extract the outputs to cleared it. The Laboratory is in a secure space that has biometric systems for the control of the access and only personnel of the laboratory can enter, and allows users to go in. It also has CCTV hours 24 surveillance.



Users must fill an application format through INEGI web site, they must load their Curriculum Vitae, document of proof of institutional affiliation and official identification. Users must be secure persons, that implies being public servant, postgraduate student, or researcher of an academic or research institutions, or personnel from international organizations. The institution of the user must sign an agreement with INEGI to provide the accreditation of the people affiliate to it. Following the practice carried by the Secure Data Service and the UK Economic and Social Research Council, INEGI signed an agreement with the National Council of Science and Technology (CONACYT) by which the researchers and students that are receiving benefits from the Council will have its accreditation and will be able to use the Laboratory, with no further requirement for accreditation. This facilitates enormously the access to those users in terms of effort and time.

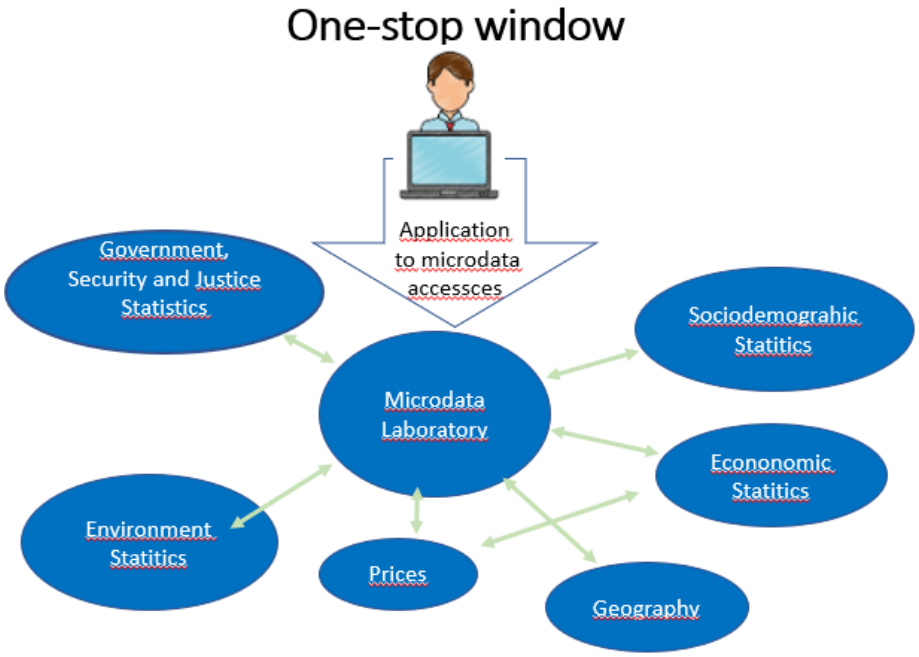
Users also need to describe the aim of its project, methodology and justify why they need access to microdata. The research project must serve public good, by supporting the definition or evaluation of public policy or else academic

research that will be published, to be publicly available. The data required must be justified in terms of the research project objective, data upon demand.

Users also need to sign a legal document of terms and conditions of usage of the Microdata Laboratory, deliver the application in original fully signed by the user and the supervisor or direct boss. Before entering their first time into the Microdata Laboratory they must undertake a training session, its content aims at creating conscientize on the importance of confidentiality and their co-responsibility in safeguarding it. We emphasize the implications of Mexico local conditions and their responsibility in safeguarding the integrity of physical and moral persons through the protection of data confidentiality. We also give them the rules of the procedure and nomenclature they must follow in other to be to handle all files involve in the attention of their research project.

3 Process of attention of microdata access service

The process of attention was defined like a star, having one-stop window with users that manages the relation with all the areas of INEGI responsible for different statistical projects, or the coordination among them when it is needed. It is them who download the microdata in the virtual desk and them who extract the output and check it. If it does present risk to breach confidentiality output is cleared and sent to the one-stop window that delivers it to the user.



4 Solutions for microdata access during COVID19 lockdown

The Microdata Laboratory was inaugurated in April 2013, at that time it had 2 cubicles with one laptop each. By January 2015 in the same space, we had to display 10 laptops to face the increased of demand for the service. On December 2016 they allocate more space to the Microdata Laboratory, so we move to a temporary location with 12 laptops. The new Microdata Laboratory was inaugurated in February 2017 with 21 laptops. Later we change the laptops used as terminals to thin clients.

The demand for the service has increased steadily, but not only in number but in complexity. In the beginning users required access to microdata of one specific statistical project and year and soon they move to request multiple years and various statistical projects. This has shown areas of opportunity, in terms of data architecture.

Year	Research projects	Remote processing	Microdata Laboratory
2011	6	6	-
2012	40	40	-
2013	125	109	16
2014	145	121	24
2015	145	90	55
2016	162	91	71
2017	146	51	95
2018	222	109	113
2019	515	276	239
2020	300	234	66
2021	219*	177	42
Total	2025	1304	721

October 26, 2021

The demand of the service kept growing, data showed clearly that having the choice between Remote Processing and Microdata Laboratory, when the conditions of the users allow it, they will take as first option, the latter.

On March 23, 2020, we were forced to close the service due to COVID19 sanitary contingency. The personal of the Microdata Laboratory move to home office using a VPN service. At first, we offered users who had Microdata Laboratory service to switch to Remote Processing, but all of them kept answering that was not a solution for them, because the Remote Processing is provided with the original databases, and they explain to us they needed to do their processing with the database they have already worked in their virtual desks. First time users did go for the only option available, and figures switch again for Remote processing as it could be expected, given the conditions of the lockdown.

Working from home office the personnel of the Microdata Laboratory could enter to all virtual desks through INEGI VPN, but only as administrators, we could see the content of the folders, but we could not use the software packages. The ICT area change the configuration of access through the VPN, so the Microdata Laboratory personnel could have the role of user and carry out processing. We developed a new service called Remote Processing in the Virtua Deks of the User. We define the whole procedure, so users would have clear expectations. Nobody was prepared for the situation we were facing, and we knew for certain that none of the users could remember the structure and names of their folders and files. We allow them to request images of their folders so they can have the names and configuration and write their do files. For the procedure we define rules of nomenclature to be able to track down any file and send them the description format for output clearance.

This new service has implied a lot of coordination, when the do file marks error we send the image to the user, and he sends back a new version of the do file. The nomenclature established facilitates the follow up of the different versions. Users have been able to continue with their analysis almost as usual, this is not available for new users, only the ones that had used the Microdata Laboratory before and had undergone training. We have offered the service to 76 users, since the beginning we established that his service will only available during the health contingency, many users that live abroad has asked us if we could continue with this modality of service.

We face another problem with the COVID lockdown with new users requesting processing with microdata from prices and agricultural census and surveys, for theses statistical project with could not offer users the microdata access through Remote Processing. So, we create a new service, we called Assisted Remote Processing in which we do allocate a virtual desk, the user sends his do file or script file and the format of the description of this output. The personnel of the Microdata Laboratory process it and then request the areas to clear the output in the standard procedure. In this case we do request users to undertake a training session by Teams in which we explain them the procedure and nomenclature. When the pandemic ends if they want to enter the Microdata Laboratory, they will need to take the full training session.

5 Conclusions

In general, the main achievements of the operation of the microdata access through Remote Processing and Microdata Laboratory is that microdata access is an institutional service before users could have access only if they had

personal connections. Nowadays all the relation is done through the web page and the institutional email microdatos@inegi.org.mx. The technical platform and process by which we offer the microdata access probe resilient to cope with the lockdown brought about by COVID19, considering local conditions that do not allow us to move to remote access. With the ICT area support we modify the services in such a way that users were able to continue with their analysis, safeguarding confidentiality.

The remarks we receive form users have made evident that we need to work more on data architecture to harness the potential of research. This year INEGI created an identifier of establishments and companies for the sole use of the microdata access services.

Also have another challenge that is the use of a unique technical platform both for users and the technical people of the areas responsible for the statistical projects, to facilitate the process of servicing users working from the Data warehouse. This will add up to security and integrity of data and can help standardized rules of data architecture. As with the original project of the Microdata Laboratory that faced organizational cultural obstacles, this new phase of having an unique technical platform will certainly face not only technical, but organizational challenges.

References

Ley del Sistema Nacional de Información Estadística y Geográfica, (2008) (Diario Oficial de la Federación México.

Ritchie, Felix. *Secure access to confidential microdata: four years of the Virtual Microdata Laboratory*, Economic & Labour Market Review May 2008, Volume 2, Issue 5, pp 29–34.

Organisation for Economic Co-operation and Development (OECD) (2014). *Microdata Access for OECD Research and Policy Analysis*, OECD: Paris France.