

**Economic and Social Council**Distr.: General
27 May 2013

English only

Economic Commission for Europe

Conference of European Statisticians

Sixty-first plenary session

Geneva, 10-12 June 2013

Item 4(b) of the provisional agenda

How should national statistical offices respond - moving from risk avoidance to risk management**Project of micro-data access of the National Institute of
Statistics and Geography of Mexico****Note by the National Institute of Statistics and Geography, Mexico****Summary*

The paper describes the current developments and remaining challenges of the National Institute of Statistics and Geography of Mexico in providing services of micro-data access for research purposes. The paper shares the experience of Mexico during a development project aimed at improving services of micro-data access.

The National Institute of Statistics and Geography has a strong commitment to supporting access to micro-data for research purposes. At the same time, it recognizes that further work is needed to develop the service. The development project that in principle sounds simple requires some re-engineering of the infrastructure of the national statistical office.

* This document was submitted late due to delayed inputs from other sources.

I. Introduction

1. The National Institute of Statistics and Geography (INEGI) of Mexico strives to harness the potential of information and communication technology (ICT) to provide access to statistical data and information. The Institute offers various tools to access aggregated statistical information through its web site, developed in 1995. The aim is to take advantage of ICT and provide as much geographical disaggregation as possible without violating confidentiality. Access to micro-data is offered, as a response to specific requests, by the different statistical areas based on their relationship with the research community. The Institute also offers a service that generates specific tables build upon user requests. This is done through the web site and the network of information centers located in all main cities of the country.

2. A new Law of the National System of Statistical and Geographic Information was published on 16 April 2008 that gave autonomy to the Institute and created a Board of Governors to regulate the National Information System. Article 100 of this Law establishes that the Institute, following best international practices, will provide access to micro-data. At the first session of the Board of Governors, one of the agreements authorized was the terms and conditions according to which the Institute will provide access to micro-data.

3. An agreement was established that access to the data of household surveys, government censuses and to the samples of the household and population censuses will be provided through file downloads from the Institute's web site. The household survey files contain raw micro-data, but confidentiality is maintained by limiting the geographic location of each unit in the random sample to municipality level.

4. This agreement also states that access to these micro-data will be provided through a research data center, but only for research purposes or to public servants that are defining or evaluating public policies. In addition, the Institute continues the regular service of processing specific tables upon request for the general public.

II. Project to develop micro-data access

5. The Board of Governors appointed a person in charge of defining and implementing a strategy to institutionalize the service of micro-data access that comprises with the following activities:

- (a) Development of a research data center;
- (b) Integration of the legal framework for the operation of the service of micro-data access;
- (c) Development of a section to the INEGI web site dedicated to micro-data access.

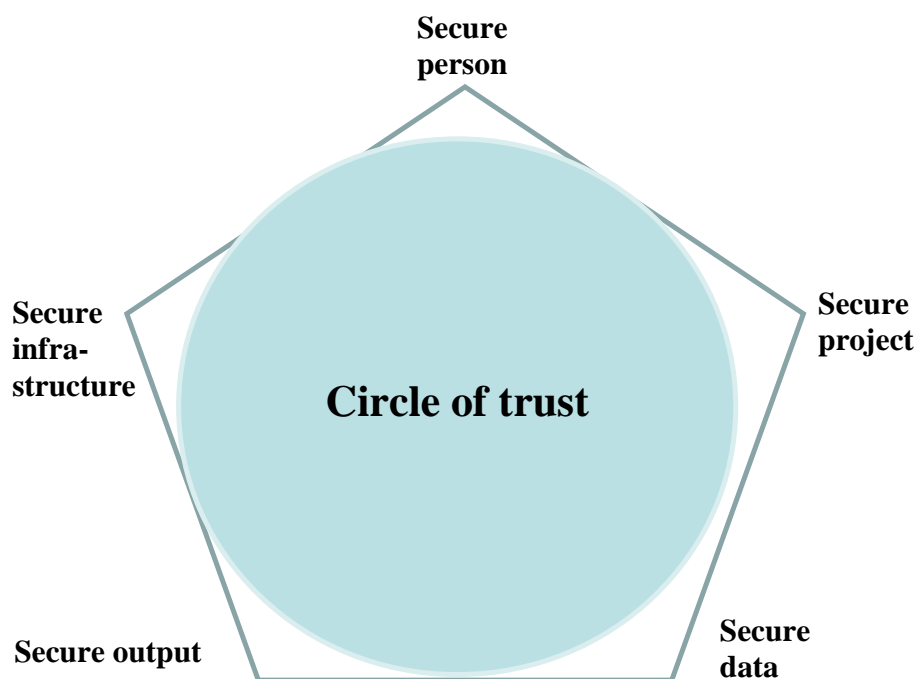
6. The person in charge of the project attended three international meetings on this subject, undertook a bibliographic research, interviewed experts, analyzed web sites of other National Statistics Offices (NSOs) and international organizations that offer access to micro-data for research purposes and visited research data centers in the United Kingdom and the United States.

7. Taking into account best international practices, INEGI decided to develop the micro-data laboratory called *Laboratorio de análisis de datos* (LAD), following the risk management model developed by the Office of National Statistics of the United Kingdom

(see Figure 1). The model comprises five elements, each of which aims at reducing risks of disclosure. The five components must be considered together to sufficiently manage the risks. The five elements are: infrastructure – both physical and technological; persons; project; data; and output. Each of these five elements must be secure in order to create a circle of trust in which an authorized user can have access to micro-data.

Figure 1

Risk management model



8. INEGI developed the technical infrastructure of the laboratory using the same type of technological platform where the Secure Data Service of the United Kingdom operates in. The platform allows the creation of a virtual space into which only authorized users with specific privileges can enter: the personnel of different statistics can access to input the data needed for the session of a specific user, or extract the outputs of the processing that the user has done to check them for confidentiality, before they are delivered to the user. The laboratory has two working rooms, each with a terminal. They are located in a secure space that provides biometric systems for access, and only the personnel of the laboratory are allowed to enter. It is also equipped with a closed circuit television (CCTV).

9. Concerning the secure persons, only public servants, postgraduate students and researchers of academic or research institutions, or personnel of international organizations can apply for access to the service of the laboratory. Each institution has to sign an agreement with INEGI to provide accreditation for the institution's affiliates. Following the practices of the Secure Data Service and the United Kingdom Economic and Social Research Council, INEGI signed, when the Laboratory was launched, an agreement with the National Council of Science and Technology (CONACYT) through which the researchers and students receiving benefits from the Council will be accredited.

10. The user will have to complete an application, provide their curriculum vitae, document of proof of institutional affiliation and official identification. In their application,

the user will also have to describe the aim of their project and the methodology as well as justify the need for access to micro-data. The research project has to serve public good or support the definition or evaluation of public policy. The data requirement has to be justified in terms of the research project's aim. The output of the processing is checked by the personnel of the statistical areas that produce the data in the first place, to ensure that confidentiality is not breached.

11. As for the legal framework, the Institute developed two norms: one that regulates the statistical information it generates and another for the National Statistical and Geographic Information System (NSGIS). It is compulsory for the Institute to provide access to the micro-data of all its statistical projects. Each of the institutions that belong to the NSGIS has to specify whether they will provide access to their micro-data. If they decide to do so, they can sign an agreement with the Institute that will allow them to use the laboratory.

12. The objective of the section dedicated to micro-data on INEGI's web site is to present to users the data sets available and the terms and conditions under which users may have access to the data.

Figure 2

A section of INEGI's web site dedicated to micro-data

The screenshot shows the INEGI website's micro-data section. At the top, there is a navigation bar with the INEGI logo and the text 'INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA'. Below the navigation bar, there are several menu items: 'Estadística', 'Geografía', 'Investigación', 'Productos y servicios', and 'Acerca del INEGI'. A search bar is also present. The main content area is titled 'Microdatos' and features a large graphic of diverse people. The content is organized into several sections:

- Censos y conteos**:
 - Censos agropecuarios
 - Censos y conteos de población y vivienda
 - Censos de gobierno *
 - Censos económicos
- Encuestas en establecimientos**:
 - Anuales
 - Mensuales
 - Especiales
- Encuestas en hogares ***:
 - Regulares
 - Especiales
 - Módulos
 - Históricas
- Registros administrativos**:
 - Económicos

On the left side, there is a text block explaining the INEGI's commitment to providing access to micro-data, followed by two bullet points: 'Directo mediante descarga gratuita de archivos de microdatos a través de esta página.' and 'Indirecto, a través de los siguientes mecanismos:'. Below these are two sub-sections: 'Laboratorio de análisis de datos para servidores públicos del Estado Mexicano, funcionarios de organismos internacionales e investigadores y estudiantes calificados.' and 'Servicio de procesamiento para el resto de la sociedad, el cual se puede gestionar a través de esta página o en los Centros de información INEGI.'

At the bottom of the page, there is a note: '* El acceso a los microdatos de estos operativos es mediante descarga del archivo, pero se puede ofrecer Laboratorio de análisis de datos si así lo requiere el usuario.' and a 'Continuar' button.

13. As part of the dedicated web site, each statistical project has its own micro-data web page (see Figure 3), except for those statistics which are accessed through download only. The micro-data pages of the statistical project have a standardized navigation system, and a common look and feel. Each page includes two subsections. The first is for all the methodological information: ideally in this subsection each project should provide:

- (a) A presentation with a short description of the main attributes of the statistical project that provides users with a rapid overview;
- (b) A methodological document;
- (c) A glossary;
- (d) A questionnaire;
- (e) A file description;
- (f) An example of the data base (with no representative data);
- (g) Catalogues.

Figure 3
Subsections of each statistical project

The screenshot shows the INEGI website interface. At the top, there is a navigation bar with the INEGI logo and the text 'INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA'. Below this, there are several menu items: 'Estadística', 'Geografía', 'Investigación', 'Productos y servicios', and 'Acerca del INEGI'. A search bar is also present. The main content area is titled 'Encuestas en Establecimientos' and features a large banner for 'Encuesta Anual de la Industria Manufacturera (EAIM) 2009 en adelante'. Underneath the banner, there is a section for 'Aspectos metodológicos' with a list of links: 'Presentación', 'Catálogos', 'Cuestionario', 'Estructura de archivos (FD)', 'Glosario', 'Metadato (DDI)', and 'Síntesis metodológica'. There is also a button labeled 'Ver todas las series disponibles'. Below this, a section titled 'Acceso a microdatos – información con mayor desagregación que la publicada' asks the user to 'Elija el tipo de servicio al cual puede tener acceso de acuerdo a su adscripción institucional'. Two options are shown: 'Laboratorio de análisis de datos' and 'Servicio de procesamiento'. The 'Laboratorio de análisis de datos' option lists user types: 'Servidores públicos de instituciones del Estado', 'Funcionarios de organismos internacionales', 'Investigadores de instituciones académicas', and 'Estudiantes de instituciones académicas de nivel superior'. The 'Servicio de procesamiento' option is 'Público en general'. Both options have a 'Continuar' button. At the bottom, there is a button for 'Otras formas de consulta'.

14. If the statistical project is already documented in the Data Documentation Initiative (DDI) standard, its DDI metadata will also be included with a hyperlink to the DDI metadata publisher called National Metadata Network. INEGI has around 141 statistical projects, and only 31 have been documented with the DDI metadata. INEGI has established an institutional commitment to document all statistical projects using the DDI metadata editor, but currently there is a lack of trained personnel to do undertake this work. This explains the need to provide methodological information in an alternative way.

15. The second subsection includes the terms and conditions by which users can access the micro-data of the statistical project. There are two ways to access micro-data: the laboratory and the compilation of specific tables. The possibility of choosing one or the other depends on the institutional affiliation of the user, as mentioned earlier.

16. Users can download the application form in this web site section, complete the form and send it by email to microdatos@inegi.org.mx

17. Even though the infrastructure of the laboratory has just been created, since the project to institutionalize the service was launched, INEGI has provided access to micro-data for research purposes to those users that require it and who fulfill the terms and conditions for remote processing. Users send their algorithms for processing the data, and the personnel of the statistical area in question processes the algorithms and checks the output for confidentiality.

III. Conclusions

18. INEGI has a strong commitment to supporting access to micro-data for research purposes – the project to institutionalize the service is still in its early stages. The Institute already recognizes that identifying the potential of micro-data sets for research purposes is complex and time consuming for the researchers, even with the DDI metadata publisher. Many researchers not only need to assess the potential of one micro-data set but also the possibility of combining variables of different micro-data sets. INEGI has identified the need to offer an “A-Z” of variables in which the user can browse through the available variables, including the metadata of the statistical project that produced the variable. This way the users can evaluate usefulness of variables from different statistical projects for their research.

19. A project that sounds simple in fact requires some re-engineering of the infrastructure of national statistical offices. It is a situation similar to that faced by the banking system in the 1990’s when changing the way banks provided their services forced them to redefine their processes, manage their data differently and undertake organizational changes. ICT allows the operation of organizational processes in ways that cannot be imagined without it.
