

UNITED NATIONS ECONOMIC
COMMISSION FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN
STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Helsinki, Finland, 5 to 7 October 2015)

Topic (v): Practicum: Case Studies and Software

Statistical data confidentiality and microdata in Albania

Gjergji Filipi* and Alma Kondi**

* INSTITUTE OF STATISTICS OF ALBANIA, Blvd Zhan D'Ark No 3 Tirana, Albania,
Gjergji.Filipi@instat.gov.al

** INSTITUTE OF STATISTICS OF ALBANIA, Blvd Zhan D'Ark No 3 Tirana, Albania,
akondi@instat.gov.al

Abstract: INSTAT increases the availability of data in web for the researchers and the public, in more easy, different and accessible format. From 2011, INSTAT has developed Instat.WebGis which gives the opportunity to all of users to select the data from different areas of Albania.

Confidentiality protection of individual, household and business data is one of the main principles of INSTAT. Best practices are set to protect confidentiality in order to guarantee that direct identification cannot be made and the risk of indirect identification is minimized following the recommendations for statistical discloser control.

This paper will describe the current situation of the statistical data confidentiality and the use of microdata and tabulated data in Albania. The legal framework, practical confidentiality experience, confidentiality declaration implemented will be also discussed.

1 Introduction

Confidentiality is one of the Fundamental Principles of Official Statistics. It is a top priority issue on the policy agenda of statistical offices and an indispensable element to maintaining the trust of respondents and thus ensuring the quality of data.¹

¹ Statistical Confidentiality and Access to Microdata, Seminar Session of the 2003 Conference of European Statistician

Confidentiality should be one of the most important personal characteristics, not only at work, but in life as well and for INSTAT all individual or business data collected are severely confidential and used exclusively for statistical purposes. Users are always seeking for more and more information from INSTAT. This increased demand from user's perspectives creates pressure between the two main basic responsibilities, to provide access to high-quality statistics and at the same time maintain confidentiality.

2 Legal Framework

Data confidentiality in the Albanian Institute of Statistics is regulated by two acts. The main act is "On Official Statistics" of 5 February 2004 adopted by Council of Ministers and was based on articles 78 and 83, point 1 of the Constitution. The principle of statistical confidentiality, the definition which statistical data are confidential, the purposes for the collection and use of statistical data and INSTAT obligations to protect confidential data are described in the law. In order for INSTAT and statistical agencies to ensure public confidence in official statistics, they follow the principles of the European Statistics Code of Practice. The organization and the implementation of the 5 Year Programme are guided by the following principles:

- Impartiality
- Reliability
- Professional independence
- Non-excessive burden on respondents
- Cost-effectiveness
- Statistical confidentiality
- Transparency

Statistical confidentiality in Official Statistics law means the protection of data directly or indirectly violation of the right to confidentiality. It implies the prevention of non-statistical utilization of the data collected and unlawful disclosure. According to this law, all statistical data collected for official statistical user has to be applied only for the preparation of statistical information. INSTAT and other statistical agencies involved in producing official statistics shall take organizational and technical measures to ensure the protection of the data fulfilled by respondent.

Another important act is The Law on PROTECTION OF PERSONAL DATA of 10 March 2008 that was based on articles 78 and 83 items 1 of the Constitution, upon a proposal of the Council of Minister. This law aims at defining the rules for the protection and legal processing of the personal data. This act regulates the management and the protection of data on natural persons.

Data Confidentiality means that controllers, processors and persons who come to know the content of the processed data while exercising their duty, shall remain under obligation of confidentiality and credibility even after the termination of their functions. These data shall not be disclosed save when otherwise provided by law. The usage and protection of microdata is not excluded in any Act in Albania, so the microdata confidentiality concept falls under the general definition of confidential statistical data.

3 Confidentiality declaration

INSTAT consider all information collected with surveys, census or administrative source confidential and protected in accordance with the process established by the law. All the information collected by the respondent is anonymized not allowing statistical units to be identified, either directly or indirectly. Statistical data are often needed for fine analysis and model building for public research like as universities, ministries and private sector also by research organizations as a part of a non-profit foundation. For the organization point of view this allow better use of collected statistical data, make better use of public money and consequently lowering the burden on statistical respondents. For each statistical institution and also for INSTAT data confidentiality is trade-off between risk avoidance and the need of statistical information for the society. According to this definition, microdata of any statistical data are confidential and have to be protected.

“Microdata are constructed from a set of records, where each record represents a data provider’s response(s) to questions included in a particular survey. Data providers range from individual respondents, households or economic entities, depending on the type of survey being carried out. A microdata file is therefore created by coding and electronically recording each survey respondent’s responses to all relevant questions.”²

In 2014, INSTAT launched a new system granting access for special authorized researches. Special authorized researches are considered scientific researchers or technical experts and/or any other party who has access to the microdata. Before the sign of the confidentiality declaration the researcher must complete an application form which includes adequate detail on the nature of the intended research, what variables are required, and the proposed outcomes of the research, such as publications, presentations, or a contribution to ongoing research.

The researcher has to sign an agreement, confidentiality declaration, with INSTAT before access is granted (Appendix 1). The confidentiality declaration sets out the requirements and responsibilities of those who have access to such information and ensures that all interested parties understand their obligations of confidentiality. The agreement, which has to be signed by the researcher before getting access to data, states that:

- To not use and to not transmit data or confidential information to unauthorized persons.
- To not use or disseminate any reserved information received during their work for any purpose that is in contrary to the aims of the work.
- To maintain the confidentiality of personal information and to keep them safe, taking all the appropriate technical and organizational measures that the files or results created based on microdata will be erased as soon as the project will be completed and such data will not be used for any other purposes.
- To preserve the image of the institution and respect the fundamental rights of individuals with regard to the principle of protection of confidential data.

² Core Principles of Confidentiality and Microdata Access in the Conference of European Statisticians (CES)

The specially authorized researches can study the complete microdata set (except for direct identifiers) stemming from all social-demographic and business surveys as well as censuses conducted by INSTAT. There is no limit on the type of analysis that can be carried out.

4 Data availability for researchers

In September 2012 INSTAT has developed a new well-functioning website with user friendly structure. The PC-Axis database enables to the user to access statistical data dynamically and downloads them in different formats. INSTAT's statistical databases have been realized with a PX-Web user interface. In the PX-Web tables are compiling by selecting classes from the variable-specific drop down menus. The table can be freely pivoted in the browser and can also be saved in various export file formats.

All statistical data produced by INSTAT are:

- Free of charge
- Available online at web in Excel format
- Available online readable from different operation system
- Can be accessed by using the PC-AXIS software in non proprietary format (CSV, XLS)
- Published on time according to the calendar of publications
- Access is not discriminatory

INSTAT is building a documentation system, metadata which is a non integrated system for the moment for all statistical activities. This system is designed to integrate the INSTAT web site, to enable the emergence of metadata corresponding to each statistical indicator published.

Most of the data derive from the administrative registers of governmental agencies, surveys and censuses conducted by INSTAT, offer information for the all users, fall into the following headlines:

- Agriculture, forestry and fishery
- Prices
- Enterprises Economic Indicators
- National Account
- Construction
- Household Consumption
- Post-Telecommunication
- Statistical Business Register
- Tourism
- Transport
- External Trade
- Education
- Living Standard
- Culture
- Population
- Environment

- Wages and Labour Cost
- Labour Market
- Health, Social Insurance and Social Protection
- Finance and Banking Statistics
- Household Income
- Energy

All statistics themes offer a comprehensive description of the purpose, history, collection, sources and methods used. Every theme page consists of five parts:

- *New* (the latest releases)
- *Figures* (preselected entries to the Statistical database)
- *Publications* (press releases, articles, etc.)
- *Methods* (definitions)
- *About this theme* (the contents of the theme and the data available)

Based on the data provided by Google Analytics for the INSTAT website it is calculated the number of page views. A page view is the amount of times visitors arrive on all individual pages of INSTAT Web site.

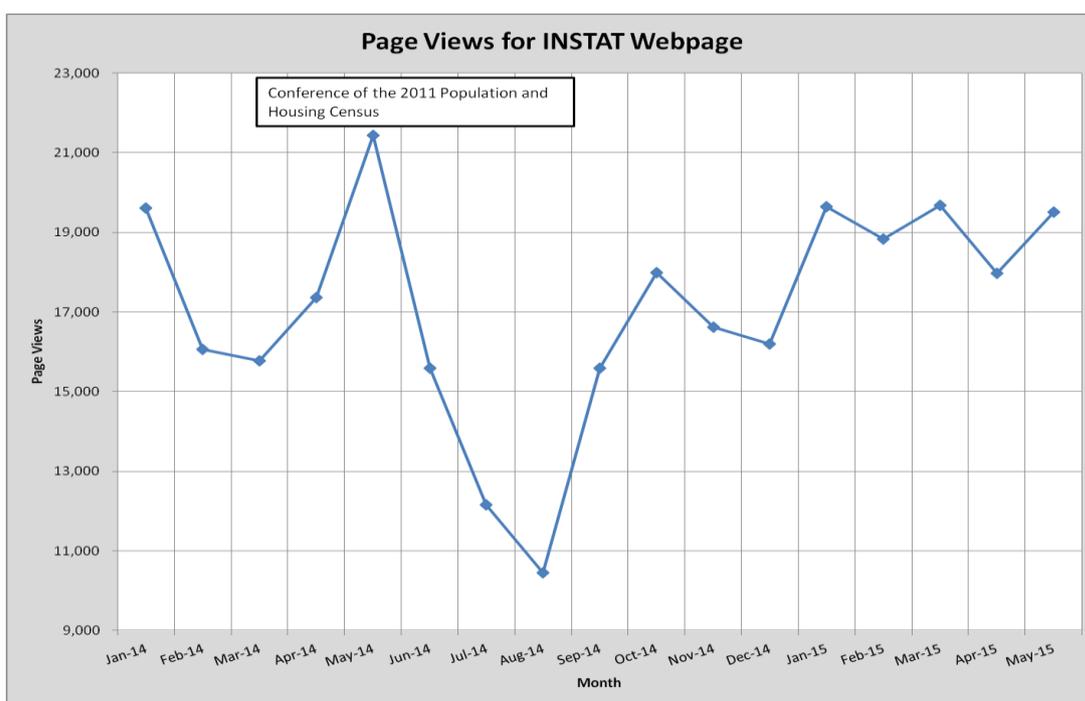


Fig 4.1 Monthly for Page Views for INSTAT Webpage

May 2014 is the most visited month during 2014-2015 with 21,432 page views. This may be connected with the Conference on the 2011 Population and Housing Census of Albania, Analysis of census data and its use. In this conference was also present instatgis, a website with all the thematic maps of Albanian census 2011. In average INSTAT webpage have 17,089 page views per month.

To avoid disclosure of confidential data on publication, INSTAT use this method:

- *Aggregation of data* - Aggregate information on entities presented in tables;
- *Cancellation* of individual data for indirect identification in statistical tables;
- *Adjustment* of the levels of details;
- *Rounding*- It is used to protect small counts in tabular data against disclosure.

4.1 Microdata

All National Statistical Institute (NSI) have to find a balance between the confidentiality protection and the increasing use of microdata. To increase use of microdata implies improved possibility of providing better data to meet the user's needs, this balance lies at the heart of official statistics which should "...provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data..."³

The new developments in technological area as hardware, software, data documentation, modern PCs now have increased the possibility to use microdata for research purposes. INSTAT, traditionally have produced aggregate statistics or to the demand users. The next development of providing value added to the data was to introduce statistical databases (microdata) that allowing the user to compile, analyze their own statistics to a large extent.

In January 2014, INSTAT have made public some microdata for all users:

- *Labour Market*: Labour Force Survey 2007 - 2013
- *Living Standards*: Living Standard Measurement Survey 2002, 2005, 2008, 2012
- *CENSUS* : Population and Housing Census 2011

Based on the data provided by Google Analytics for the INSTAT website it is calculated the number of microdata downloads for each month using Page Views. In this case user cannot see the content of microdata before downloading it. For the scope of calculating the number of downloads each Page Views is associated with downloads.

³ UN Fundamental Principles of Official Statistics, Article 1, UN Statistical Commission 1994.

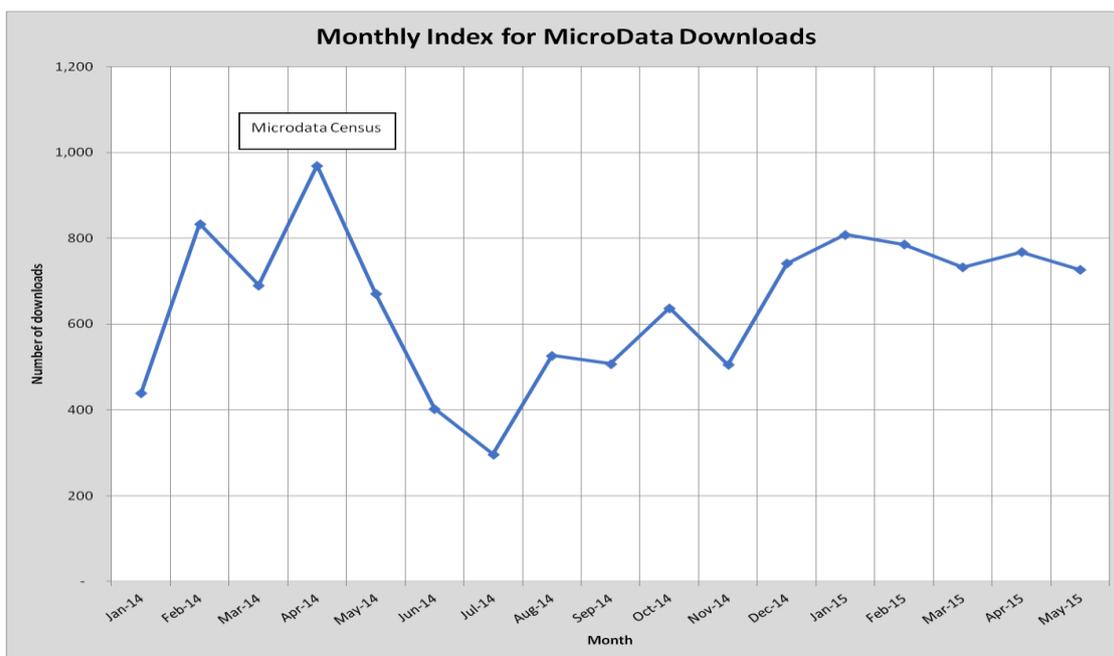


Fig 4.2 Monthly Downloads for Microdata

In average the number of downloads of microdata per months is 650 downloads. The month with the highest number of downloads is April 2014 with 970 downloads. This may be explained by making available to the public the 3 % of microdata of the Population and Housing Census 2011 database.

To avoid disclosure of confidential statistical data on microdata, INSTAT use:

- *Anonymisation* – deleting direct identification data;
- *Recoding* variable for indirect identification
- *Sampling* – make a sample so a small proportion of the original data set as a microdata file;

4.2 WebGis

A web page (WEBGIS <http://instatgis.gov.al>) is now available in the website of INSTAT from May 2014. WEBGIS is an innovation in the frame of increasing the role of geo-statistics and gives the possibility of comparing the geographical distribution of statistical data. The thematic maps of Albanian census 2011 data are now available for all interested users about statistics. These data are displayed based on the indicators of census questionnaire and are classified into 3 categories, which are: population, education and employment and housing and living conditions in different levels.

The structure of WebGIS application is very simple and very easy to understand. Users can navigate through different panels:

- The main view of the “map” with the Albanian boundary focused on it.

- The part of “geography and indicators”, which allow users to choose one of the three levels to map the indicator they want.
- The “classification” method allow the user to manipulate and analyse the data in any way they want: the classification method (manual, natural-breaks, equal-interval, quintile, geometrical interval) , the number of class breaks , the transparency of layer of thematic map, the colour ramps
- The “map description” where the users can find the description of the selected indicator.
- The “print, share and language”, which allow the users to select the language (in Albanian or English language), to print the map, to share the map in different social networks.

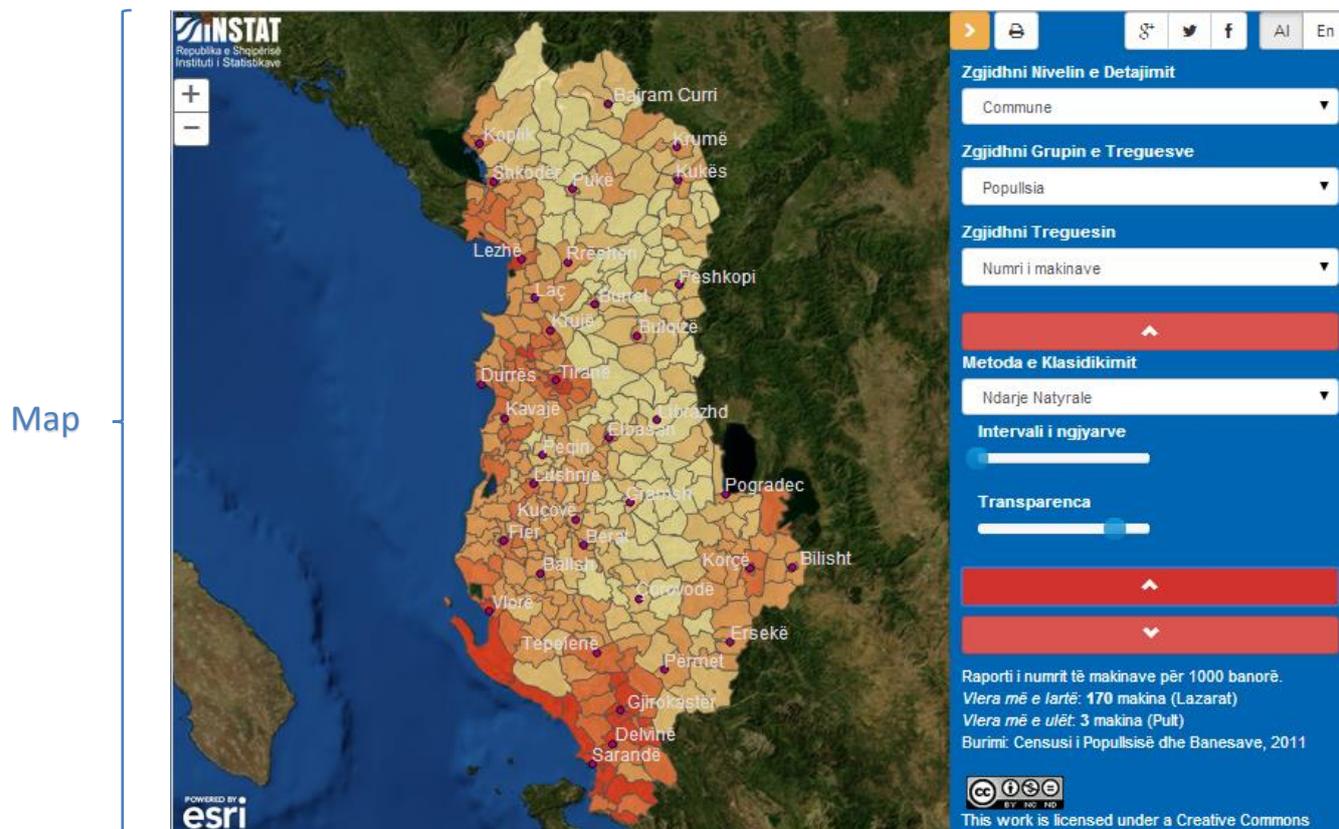


Fig.4.3 WebGIS interface

The maps can be printed from the website by the users according to the model created by INSTAT. In the future the WebGIS will be populated with other statistical indicators.

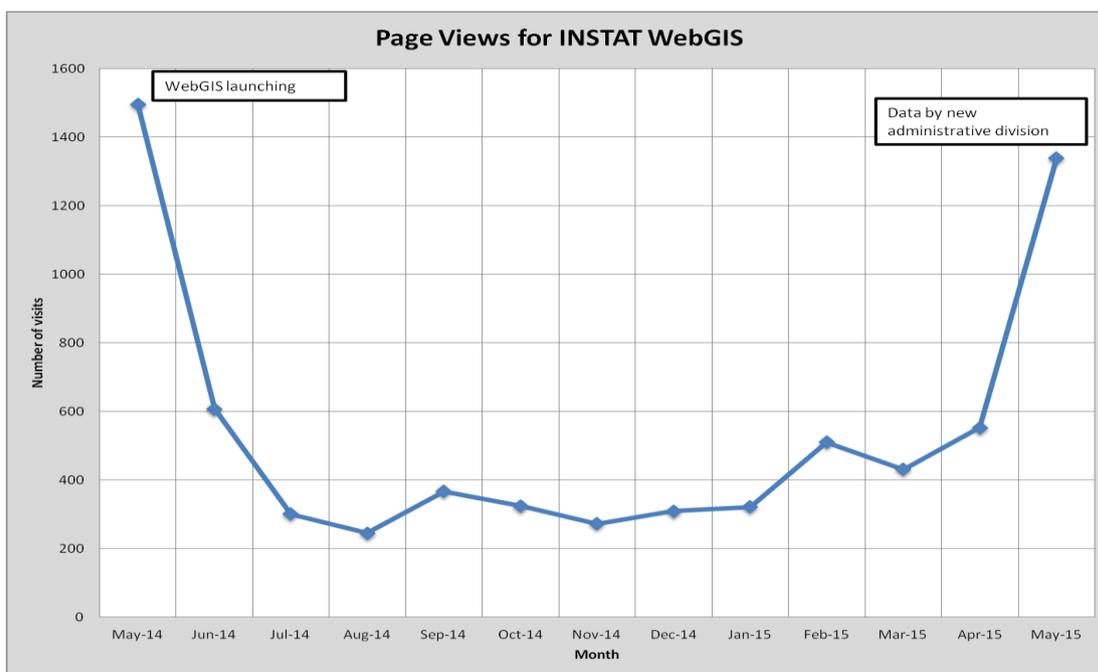


Fig 4.4 Monthly Page Views for INSTAT WebGIS

May 2014 is the most visited month during the period May 2014-May 2015 with 21,496 page views. In May 2015, INSTAT has publish the data of Population and Housing Census 2011 based on the new territorial-administrative, with 61 municipalities. The number of page views for INSTAT WebGis is 1,338. In average INSTAT WebGis have 544 page views per month.

5 Conclusion

INSTAT collects data from individuals and businesses for the scope of producing official statistics. After that its work is concentrating on manipulation, analyzing, presenting and disseminating statistical products. All these activities must have a legal and ethical base in order to ensure the privacy of those providing the information. In this way the individuals and organisations cannot be identified in a disseminated dataset.

There is a trade off between confidentiality and transparency. Users are more and more interested in disaggregated data. The problem is at what level can INSTAT produce data that either fulfil user needs or protect data confidentiality. Regarding confidentiality INSTAT base its activity on a legal framework. On the other hand a lot of efforts are made in the direction of improving data dissemination. INSTAT web site it's a powerful tool of disseminating statistical products. Microdata on the other hand are an approach of improving transparency to the public. In this way users capable of using statistical software do their own analysis for their research purposes. Pc-Axis table allow users to do their own tables. This requires online data validation and rules which control for data disclosures. New tools for data dissemination bring the challenge of using more and more sophisticated software for controlling and protecting direct identification of respondents.

References

Core Principles of Confidentiality and Microdata Access in the Conference of European Statisticians (CES)

Law No. 8517, dated 22.07.1999 "On the Protection of Personal Data"

Law No. 9180, date 5.2.2004 "On official statistics"

Statistical Confidentiality and Access to Microdata, Seminar Session of the 2003 Conference of European Statistician

UN Fundamental Principles of Official Statistics, Article 1, UN Statistical Commission 1994.

Appendix 1

DECLARATION OF CONFIDENTIALITY PROJECT: Updating the poverty maps in Albania with 2011 Census data and 2012 LSMS data

Object

This declaration is addressed to INSTAT from: scientific researchers or technical experts and/or any other party who has access to the file with variables created on 2011 Census data, which will serve to update the poverty maps in combination with data from LSMS 2012, in the framework of cooperation with the World Bank.

Purpose

This declaration shall be signed by all individuals who have access to the file with variables created based on 2011 Census data or unpublished data. The declaration sets out the requirements and responsibilities of those who have access to such information and ensures that all interested parties understand their obligations of confidentiality.

Field of application

This declaration of confidentiality applies to all confidential data and information known during the work for updating the poverty maps in Albania with 2011 Census data and 2012 LSMS data. The relevant provisions apply even after the work and relevant analysis is completed, regardless of whether the declarant is still working or not with INSTAT.

Declaration of Confidentiality

1. Through this declaration, I undertake not to use and not to transmit data or confidential information to unauthorized persons, about or obtained from 2011 Census data and 2012 LSMS data, unless expressly authorized from Director General of INSTAT, or when required by law. I understand that this obligation applies during the terms of operating data, as well as after its completion.
2. I understand that the use and disclosure of up mentioned data are subject to the Law no. 9180, date 5.2.2004 "On official statistics" as amended and to the

Law nr. 9987, date 10.03.2008 on "Protection of Personal Data" as amended.
I will not use or disseminate any reserved information I receive during my work for any purpose that is in contrary to the aims of this work.

3. I am fully aware that I am required to maintain the confidentiality of personal information and to keep them safe, taking all the appropriate technical and organizational measures that the file with variables created based on 2011 Census data or results is to be erased as soon as the project will be completed and such data will not be used for any other purposes other than the up mentioned project.
4. I take full responsibility if found to have violated the guidelines regarding confidentiality of data or in case of not preserving them, immediate actions are applied. I understand this action as a need to preserve the image of the institution and respect the fundamental rights of individuals with regard to the principle of protection of confidential data.

Name: _____

Position: _____

Signature: _____

Date: _____