

# Enhancing Quality Practices at the Brazilian Institute of Geography and Statistics

Maria Luiza Barcellos Zacharias

*Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil; maria.zacharias@ibge.gov.br*

## Abstract

In the last years, the IBGE, the Brazilian Institute of Geography and Statistics, has been focusing on different aspects of the quality of statistical production, while taking into consideration the best practices and the principles of official statistics. Many significant efforts have been implemented in order to improve the accuracy, comparability, accessibility and transparency of the statistical production process. The edition of the IBGE Code of Practice, the release of a set of Guidelines and Protocols related to the dissemination of statistics, the increasing adoption of international standards and frameworks, like DDI, SDMX and GSBPM, are some examples of these efforts. In addition, one of the major projects in progress to improve quality and efficiency is the redesign of the corporate statistical metadata system. The intent of the metadata modernization is to establish common semantic structures, robustness of systems and harmonization of the survey questions and of the variable names, concepts and definitions, in a user friendly environment. Furthermore, the metadata structure is being developed in alignment with international standards, in particular, ISO/IEC 11179 and the Data Documentation Initiative – DDI standard. The aim of this paper is to present the progress and development of some quality management practices implemented recently, as well as the initial steps towards the modernization of the metadata system.

**Keywords:** Statistical quality, Metadata, International Standards

## 1. Introduction

Several developments for improving statistical quality management have been continuously implemented at the Brazilian Institute of Geography and Statistics (IBGE) in the last 8 years. The paper will present some of the initiatives that have been taken and others that are in progress in order to improve the quality of the information produced by the IBGE.

## **2. The IBGE Code of Practice**

In 2013, the IBGE made available its Code of Practice, disseminating a set of guidelines, principles and indicators on the best practices adopted by the Institute in the production of official statistics, taking as reference the Statistics Code of Practice for Latin America and the Caribbean. It sets out 17 key principles and 80 indicators, concerning the institutional environment and coordination; statistical processes and statistical products, that can be used for monitoring the implementation of the Code. It is also seen as an important step towards improving the governance and consistency of official statistics in Brazil, because it aims not only at fostering the quality statistics debate in the several agencies of the National System of Statistics - NSS, but also at providing information to the creation of a more comprehensive version of this document, embracing the statistical production of the whole system. Specifically, the objective is to introduce a common understanding of quality among all the producers of statistics in Brazil and also to align national practices to international standards. To accomplish this future edition of the NSS Code of Practice, the IBGE is promoting discussion forums with other institutions responsible for the production of official statistics, firstly at the federal government level. Soon after the release of this publication, a set of guidelines and protocols have been disseminated to provide guidance and support in compliance with the principles established by the Code of Practice.

## **3. Guidance and Protocols**

One of the IBGE's main objectives is to ensure that the information produced is made available on an impartial basis to honour citizens' entitlement to public information, according to the first of the Fundamental Principles of Official Statistics. In order to disseminate widely this commitment among its technicians and Brazilian citizens, in October 2014 the IBGE published on its website the "Protocol on Impartial Access to the IBGE Data", containing guidelines related to both oral and written dissemination of official statistics, ensuring the compliance with the principles of equal access and impartiality of information.

In April 2015, the IBGE established other important protocol, the "Guideline on how to deal with Publication Errors", in order to ensure uniform error management. It is important

to disseminate standard procedures of dealing with publication errors, where incorrect data are published, despite all the quality assurance measures taken. In accordance with the IBGE Code of Practice, in its principle 15 - Timeliness and Punctuality, indicator 15.4: "errors discovered in published statistics must be corrected at the earliest possible date and publicised". In order to maintain confidence in official statistics in case of errors, it is important that the IBGE reacts reasonably and in a uniform and transparent way. Therefore, the whole procedure of dealing with publication errors in the IBGE is described in this guideline, which comprises an error typology and the response to each type of publication errors, considering the different forms of disseminating statistical results. This protocol was heavily inspired by a similar guideline made available by the Federal Statistical Office of Germany (DESTATIS, 2013).

In addition, two new protocols have been recently approved, reinforcing the aim of establishing standardized procedures: one of them establishes the IBGE procedures concerning the revision of statistical data, and the other one addresses good practices for dealing with misuse of statistical data.

Regarding the first protocol on revision of data, it describes the procedures in relation to planning, time of implementation and purpose of changes to statistical data. The reasons for revising data may be, for instance, the inclusion of an additional data source or the transition to a new base period. At the IBGE, each statistical operation adopts its own data revision policy according to the type of information it produces. So, for example, the consumer price index used as the official inflation index is not subject to revisions, for legal reasons. On the other hand, some statistical business surveys revise their initial estimates because further data becomes available, and new, more complete or additional data that was not available at the time of the first release of statistics are included. The revision policy for those outputs that are subject to scheduled and non-scheduled revisions is publicized at the IBGE website in order to be made clear to all users.

In relation to the second protocol mentioned, the Fundamental Principle 4 establishes that the statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics, which is also referred in the IBGE Code of Practice, in Principle 1 – Institutional Independence, indicator 1.7, "The statistical authorities of the IBGE, when appropriate, should comment publicly on statistical issues or advise on public comments, including

criticisms and misuses of statistics.” Types of misuse include: discarding unfavorable data, using loaded questions, overgeneralization, biased samples, misreporting or misunderstanding of estimated error, false causality, proof of the null hypothesis, data dredging, data manipulation, and class fallacies. The protocol on misuse of statistical data provides good practices to be adopted in case of misuse, a wrong interpretation or even an overreaction from the media, society or government is detected, related to the statistical information released.

## **5. Redesign of the Statistical Metadata Base**

The Metadata base (METABD) is the corporate repository of metadata for each of the IBGE’s current 62 surveys and statistical programs. It contains another 89 records for surveys that have been discontinued or with no publicly disseminated data, for historical purposes. The content of the METABD has been selected to suit its primary purpose, which is to provide users with information needed to interpret the statistical data disseminated by the IBGE. The type of information provided covers mainly the methods used to produce the data published from surveys and statistical programs. The METABD is organized around the survey entity, containing both reference and structural metadata. Reference metadata describe key characteristics, as methodology, reference period, geographic coverage, dissemination forms, etc. of the survey. Structural metadata refer to variables and their definitions, and related classifications.

Recently, the IBGE has undergone a wide redesign of METABD. The intent of the metadata modernization initiative was to establish common semantic structures, robustness of systems and harmonization of survey variables, questions, concepts and definition, in a user friendly environment. Furthermore, the metadata structure is being developed in alignment with international standards, in particular, ISO/IEC 11179 for defining variables names, and the Data Documentation Initiative – DDI standard, particularly the Codebook 2.5, to support the definition of the contextual and descriptive metadata that should be used to describe the surveys, variables and files. The new metadata system is compliant with DDI standard in such a way that all metadata information can be automatically converted in DDI files, generating XML documents in a standard text format that facilitates data communication between organizations and software systems and improve the quality of

statistical documentation provided to users of data, by means of a structured framework for organizing and disseminating information on content and structure of statistical information.

The work currently in progress involves the registration, from scratch, of the whole set of variables of the surveys that are conducted by the IBGE. It is expected to move the agency towards greater quality and coherence in terminology used for metadata by means of harmonization. A Metadata Harmonization working group has been created at the IBGE in order to standardize variable names, to harmonize survey questions, concepts and definitions to ensure that the quality of official statistics meets user requirements, making it possible to compare data from different sources with confidence and to merge and match data more easily. It is possible to apply the variable naming structure in a standardized way because of automation resulting in improved efficiency, reliability, maximizing re-use and minimizing human error despite the volume of variables. The naming of the variables criteria was defined jointly by the METABD and the surveys staff based on the ISO 11179-1 naming convention, using the three elements to create the name of the variables – PROPERTY + STATISTICAL UNIT + REPRESENTATION FORM. To register a variable in the METABD, these three elements are informed, selecting properties that have already been included or including a new one, and selecting the statistical unit and representation forms using selection boxes. The system then automatically generates the variable name and, to avoid human intervention and typing errors. A broader concept of the variable also has to be provided in order to create the variable in the METABD. Afterwards, other metadata attributes are filled, related to the variable associated to that specific survey, like its operational definition, question text and number, rules of derivations, etc. Once a variable is created and associated to a specific survey, it is immediately available to be reused by any other survey, making only the adjustments needed to adequate the specific metadata for that survey. So, if there is a metadata registry with specific metadata already in it, we need not to develop another but rather develop the potential of the existing registry, making small changes needed, maximizing the re-use of information. The loading of the variable metadata information in METABD is made by the survey teams and is subjected to a registration process controlling authorization and further validation of the information and conformity to standards, by the METABD staff.

The harmonization strategy and the standardization of the variables names will bring many benefits to both producers and users of statistics, such as:

- Better comparability for cross-survey analysis and integration by making statistics comparable across time and across different sources, thereby creating an environment in which multiple data sources can be integrated;
- Improved data quality by reducing potential misunderstanding or confusion caused by the use of different questions from the same topic area;
- Improved ease of use and comparability – for data users who require multiple datasets, having multiple concepts with multiple meanings could not only be arduous and time consuming to handle, but increase the possibility of mistakes or make results incomparable.

It must be mentioned that a significant part of the conceptualization and the definition of the variable module of METABD was inspired by the work done by Statistics Canada, related to the Integrated Metadatabase (IMDB).

## **6. Governance and National Statistical Coordination**

The Fundamental Principle 8 establishes that coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system. Nonetheless, according to Global Review (UNSD, 2013), this principle is among the least implemented of the Fundamental Principles of Official Statistics. In Brazil, there is a legal framework that establishes a National System of Statistical Information and states that the responsibility of regulating and coordinating that System will be the charge of the National Statistical Office, i.e., the IBGE. Although this coordination does not take place comprehensively, the IBGE plays an important role in providing technical and methodological guidance on technical matters of statistics products and services to specific demands from other statistical producers.

In order to promote the effective coordination and governance of the National Statistical System, the Social Statistics Committee was established, in December 2007, as a forum, coordinated by the IBGE, consisting of the main producers of social statistics of the federal government aiming at the integration of the various databases derived from surveys and

administrative records. The following guidelines and rules were set for guiding the statistical work of the members of the Committee:

- To ensure and to improve the comparability and accessibility of information;
- To encourage the adoption of concepts, classifications and methods for processing, archiving and presenting official statistical information;
- To fill in information gaps and expand the integration of data sources;
- To organize the demand for information and coordinate their production and use;
- To regulate and to facilitate access by various government agencies to existing databases, ensuring transparency, visibility and confidentiality;
- To adopt the Fundamental Principles of Official Statistics.

Another important initiative towards the coordination of the National Statistical Systems refers to the use of a common structure of metadata among the main producers. The IBGE has been promoting short courses, implementing a training plan in order to develop skills and professional expertise on documentation of statistical operations and administrative records, using the standard adopted by the IBGE METABD, in compliance with DDI standard. The plan is to congregate all data and metadata bases of the members of the National Statistical System under a common IT and methodological infrastructure, in order to organize, regulate and harmonize the statistical production of the country. Specific standards should be established to preserve, capture, analysis and exchange of information by electronic means between the IBGE and the main producers of the national statistical system. The development of a national integrated information system, similar to a National Data Archive, is now part of the federal government quinquennial plan, a challenge that will require strong interagency articulation while aiming at the efficiency and the benefits of having a national statistical system that upholds quality, relevance, objectivity and accessibility of the statistical information.

## **7. Modernization project - future applications of the GSBPM**

The IBGE has decided to adopt GSBPM 5.0 as a reference model for use through the agency, as a first step towards the modernization of the agency, following the international work coordinated by UNECE to promote standards-based modernization, Modernstats. There are three different initiatives being planned to improve the data quality of the IBGE products and the efficiency of its statistical programs, involving the use of GSBPM.

The first application of GSBPM will be to ensure data quality, as a framework for reviewing several statistical programs for their quality assurance practices and identifying which sub-processes have a greater risk of errors in quality. At the IBGE, every statistical program, as well as all infrastructure and operations programs, has their own data quality checks and quality assurance practices. However, they are not standardized nor organized in a structured way, what can lead to the occurrence of recurring errors. As with any large and complex organisation, problems with processes do arise and the IBGE has suffered errors in their data in the recent past with varying degrees of impact on the public domain. Most errors are detected in-house before publication, while other ones have only been discovered after release, resulting in re-issue of statistical output, leading to emerge credibility issues. As a result of these errors situations, the IBGE decided to put forward better quality management practices through the development and use of the risk mitigation strategy known as quality gates, inspired by the work developed by the Australian Bureau of Statistics, ABS. Quality Gates are a quality management process designed to improve the early detection of errors or flaws in production processes. For detection of errors in statistical agencies, GSBPM can be used as a guide to map the activities of the processes along the statistical production, in order to identify where checkpoints must be placed throughout a process to identify errors or problems early (ABS, 2010).

The second application of GSBPM aims at the efficiency of the processes. The IBGE is facing the challenge of maintaining the quality of its information products while using increasingly less resources. In this context, the agency is looking at optimizing and standardizing the processes involved in the production of statistical output, and there is a work in progress to develop an Integrated Business Statistics Program, using GSBPM as reference framework, as a way to achieve these objectives in business statistics, as a pilot study. The objective is to use GSBPM to align services, to align underlying information management with business strategies, to increase efficiency of the statistical processes, with surveys sharing common sampling, collection and processing methodologies, as well as common tools to edit, correct and analyse data, enabled by the description of the business processes using a common terminology.

Finally, the third planned use of GSBPM is for the description of the metadata of the processes and its subprocesses. The IBGE is working on the elaboration of metadata that

embraces the whole statistical process, which requires the mapping and the description of all the processes, subprocesses, with the identification of their inputs and outputs, as part of the wide redesign of the current METABD, in order to move it towards an integrated metadata system. In this way, the new statistical metadata strategy is to transform the current metadatabase into a more comprehensive system, which includes not only the statistical reference and structure metadata, but also the processes metadata, the quality metadata and the statistical quality management, linked to a documentation repository to store documents and materials produced along the statistical production processes, for all surveys. Considering that the current metadata system (METABD) is already linked to the statistical collection, i.e., the micro-database files, and also with the aggregate statistical data dissemination system, this new integrated system would be the mainspring of the statistical data elaboration process and a strategic statistical information repository, consolidating a metadata-driven approach for the statistical production, as recommended by UNECE Modernstats project.

## **7. Conclusion**

The Brazilian Institute of Geography and Statistics is continuously implementing initiatives to enhance the management of statistical quality, in order to improve the accuracy, comparability, accessibility and transparency of the statistical production process. Although much has been done recently, a lot of significant efforts are currently in progress and still being implemented to meet the relevant quality standards required, to modernize the processes of statistical production, by the increasing use of international standards, and to improve the consistency and efficiency of the national statistical system, by reinforcing the IBGE's coordination role.

## References:

ABS - Australian Bureau of Statistics (2010), Quality Management of Statistical Processes Using Quality Gates. Information Paper 1540.0.

DESTATIS - Federal Statistical Office of Germany (2013), How to deal with Publication Errors – Guideline.

DUPRIEZ, Olivier; BOYKO, Ernie. (2010), Dissemination of Microdata Files. Formulating Policies and Procedures. International Household Survey Network, IHSN Working Paper No 005.

IBGE. (2014), Conjunto mínimo de indicadores padrão de qualidade a ser aplicado no MERCOSUL. Textos para discussão nº 52.

IHSN. (2012), IHSN technical note on metadata standards – DRAFT. Available at [http://www.ihsn.org/home/sites/default/files/resources/DDI\\_SDMX\\_IHSN\\_DRAFT.pdf](http://www.ihsn.org/home/sites/default/files/resources/DDI_SDMX_IHSN_DRAFT.pdf)

ISO/IEC 11179, Information Technology -- Metadata registries (MDR). Available at <http://metadata-stds.org/11179/>

LEE, J., MARTINEAU, S. (2015), Developing Metadata Standards in an Integration Project at Statistics Canada. Paper presented at the Workshop of the Modernisation Committee on Standards: International Collaboration for Standards-Based Modernisation.

JOHANIS, P.; GILLMAN, D.W. (2006), Metadata standards and their support of data management needs. Paper presented at the Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS). Geneva.

OECD Principles and Guidelines for Access to Research Data from Public Funding. (2007) Organisation for Economic Cooperation and Development (OECD). Available at <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

UNSD, United Nations Statistics Division (2013): Implementation of the Fundamental Principles of Official Statistics Background Document. Available at <http://unstats.un.org/unsd/statcom/doc13/BG-FP.pdf>

UNSD, United Nations Statistics Division (2015). United Nations Fundamental Principles of Official Statistics – Implementation guidelines,. Available at [http://unstats.un.org/unsd/dnss/gp/Implementation\\_Guidelines\\_FINAL\\_without\\_edit.pdf](http://unstats.un.org/unsd/dnss/gp/Implementation_Guidelines_FINAL_without_edit.pdf)