Modernization of statistical production through the implementation of the GSBPM

Grace Andrea Torres
Víctor Andrés Arévalo
José Anderson Castañeda
Regulation, Planning, Standardization and Normalization Division (DIRPEN)
National Administrative Department of Statistics (DANE) – Colombia

Abstract: This paper explains the experience of the National Administrative Department of Statistics, DANE with respect to the modernization of statistical production based on the analysis of the GSBPM standard version 5.0. The purpose of the project is to show that statistical operations produced by DANE can be improved through the industrialization of sub-processes and activities based on the GSBPM model. This project will be developed in three stages. The first stage compares the GSBPM model with DANE's statistical production model in order to identify the technical, operational and management-related issues in DANE's model. The second one develops mechanisms for implementation, expressed in a work plan on the medium term; the third stage refers to the application and evaluation of this work plan. This document will focus on the first stage of the project where the opportunities for industrialization were identified. These issues were assessed by the thematic experts responsible for the sub-processes. Finally, the document will show the opportunities identified in the process of modernization of statistical production related to the integration of model components.

Key words: Statistical Modernization, Generic Statistical Business Process Model (GSBPM), Statistical Production Model, and DANE.

1. INTRODUCTION

The world is experiencing a data revolution. As information technology is rapidly expanding everywhere, tremendous amounts of information are being automatically recorded and stored, leading to the question of whether National Statistical Organizations (hereinafter NSOs) can use it to produce official statistics. This data revolution also means new expectations on behalf of data users: Increasingly governments, development professionals, media, or general public call for statistical data available in a faster way, at a more detailed level of disaggregation, in a more integrated manner by linking data over a wide range of topics (ESCAP, 2016).

In order to meet these needs, the NSOs must go through a transformation process, both in the way the National Statistical Systems operate and the way in which official statistics are produced. In this regard, modernization requires these necessary transformations.

According to the Economic and Social Commission for Asia and the Pacific of the United Nations (ESCAP), modernization is understood as the "transformation in the way that the National Institutes of Statistics work, incorporating common generic processes, common tools and common methodologies "(Javorsek, n.d. h), which involves both recognizing that th statistics are produced in a similar way and increasing the capacity and flexibility to adapt to change. Understood in this way, modernization requires two drivers:

a) The creation of a more adaptive and cost-effective information management environment through stronger collaboration within the statistical offices, among the entities that make up the national statistical systems and the countries of the region, which entails achieving processes that are more agile and of better quality that require less resources to operate and maintain.

b) The improvement of information systems to address the evolution field of statistics and maintain relevance across developed and developing countries.

For this reason, the international statistical community has developed some standards in order to meet these challenges, such as the Generic Activity Model for Statistical Organizations (GAMSO) (UNECE, GAMSO V1.0, 2016); the Generic Statistical Business Process Model (GSBPM) (UNECE, GSBPM V5.0, 2016); the Generic Statistical Information Model (GSIM) (UNECE, GSIM, 2016); the Common Statistical Production Architecture (CSPA) (UNECE, CSPA v1.5, 2016); the Standard Data and Metadata Exchange (SDMX) (Community, 2016) and Data Documentation Initiatives (DDI) (World, 2016).

The National Administrative Department of Statistics (DANE)´s modernization strategy as an entity producing statistics lies on the strengthening of its statistical production model (hereinafter SPM), which essentially makes up the basis of its network of processes and its value chain. This strengthening aims to improve the integration, interoperability and comparability of the official statistics generated by DANE with respect to those produced both by entities belonging to the Colombian National Statistical System (hereinafter NSS) as well as those generated by other national statistical offices and other international organizations.

A method to realize this *modernization strategy* is to review the DANE's model based on the Generic Statistical Business Process Model[1] –GSBPM-, which considers all the activities covered by the producers of official statistics that culminate in the dissemination of results and is valid for national and international statistical organizations, since it is based on the purpose of building an "international generic standard", taking advantage of the maturity and experience achieved by official entities responsible for the statistical production of different countries[2]. The strengthening of the SPM is required to enhance, among other things, three attributes of the system: adaptability to change; effectiveness (performance); and confidence (credibility).

---

[1] The GSBPM model v.5.0 is being updated and complemented with the so-called GAMSO - Generic Activity Model for Statistical Organizations, published by September 2015 by UNECE.
[2] "Towards a Generic Statistical Business Process Model". Steven Vale. UNECE

In this regard, as part of efforts to consolidate the modernization strategies in DANE, the Regulation, Planning, Standardization and Normalization Division (DIRPEN, for its acronym in Spanish) has been carrying out a *modernization project* through the implementation of the GSBPM, which consists of three stages: in the first one, DANE´s SPM was compared with the GSBPM model in order to identify specific points for improvement per phase, and general points for improvement in the DANE model and, in turn, opportunities for industrialization, which were discussed with the officials responsible for the sub-processes; the second establishes mechanisms for the implementation of these opportunities, expressed in a work plan in the medium term; the last stage will be the implementation and evaluation of the plan.

This paper will focus on the first of these stages and will consist of four parts. The first part presents the standards covered by DANE for the modernization of its statistical production, including the GSBPM. The second part presents the methodology that was raised from DIRPEN to analyze the GSBPM as a viable standard to modernize the statistical production. This led the team to a mapping of the phases and sub-processes of both DANE and the GSBPM model, which allowed finding divergent and convergent points. The third part discusses the findings of such mapping in terms of the areas for improvement. The fourth part explains the identified opportunities. Finally, we provide some conclusions with respect to the adaptation process and the steps outlined in DANE in order to continue the statistical production modernization process.

2. STANDARDS FOR MODERNIZATION INCLUDING GSBPM

In a joint effort between Statistics New Zealand, Sweden, Australia, Norway, the Netherlands, the Economic Commission for Europe of the United Nations (UNECE), the Statistical Office of the European Union (Eurostat) and the Organization for Cooperation and Economic Development (OECD), various projects have been developed pertaining to the collection of best practices for the development of a statistical process, which gave rise to the Generic Statistical Business Process Model (GSBPM) as a standard for statistical production. This model divides the process into a series of phases or stages, with common elements such as the specification of user requirements, the processes to design and build the operation as well as the collection and processing of data and evaluation phase.

A basic principle of this model is to ensure quality in all these phases. It also takes steps to ensure that the design is properly made for the implementation and execution of the operation, making clear that a good design is not enough; it shall be accompanied by a proper implementation and a comprehensive evaluation process so as to verify whether the statistical program meets its objectives.

The GSBPM is applicable to all the activities carried out in the production of official statistics, in order to integrate and interoperate the various phases of statistical production. The model refers to the phases of the statistical process of a statistical office (also known as the statistical value chain or statistical cycle) and provides generic terms to describe them. The GSBPM functionality model is summarized in the following aspects:

• It becomes a general benchmark of a statistical process without being restrictive, too abstract or theoretical.

• In every sub-process of the model it allows identifying the necessary inputs, the expected outputs, purposes, person responsible, methodological documentation, technical and human requirements as well as feedback mechanisms.

• It allows the comparison of reference processes within and between organizations.

• It promotes synergy between the model processes and the implementation of standards such as GSIM, CSPA, DDI and SDMX; furthermore, it is the basis for the implementation of the GAMSO standard.

2.1. Operational structure of the GSBPM

Although the statistical process focuses on the collection and processing of data to produce statistical results, the GSBPM can also be applied to the redesign of a statistical operation or to derived statistics. In these cases, the input data are those published previously, which are then processed and analyzed to produce results. In these situations, it is likely that multiple sub-processes and possibly some phases (particularly the early ones) are not documented or not delved into from a methodological standpoint.

In addition to being applicable to various types of statistical operations, the GSBPM can also be applied to the development and maintenance of statistical records, where inputs are similar to the statistical production inputs (although usually the statistical production have a greater focus on administrative data) and outputs are frames or other data extractions, which are then used as inputs for other processes.

From the above, it is concluded that the GSBPM is designed to be applied regardless of the data source, hence it can be used for the description and evaluation of the quality of statistical operations such as censuses, sample surveys or based on administrative records, or even those resulting from the combination of other sources of information.

The GSBPM, version 5, comprises four levels:

• Level 0: The statistical production process;

• Level 1: The eight phases of the statistical production process;

• Level 2: The sub-processes within each process;

• Level 3: A description of the sub-processes.

The structure of the GSBPM model specifies eight phases that in turn consist of 44 sub-processes.

It is noteworthy that each sub-process must identify input, output, purpose, point-person, guidelines (manuals and documentation), facilitators (persons and systems) and feedback. The GSIM[3] is behind the GSBPM model, which allows identifying the inputs and outputs of each of the sub-processes, creating objects of information. The Common Statistical Production Architecture or CSPA is also included, which corresponds to the technological architecture, therefore, is considered relevant to link the GSBPM and GSIM standards.

Also, in order to strengthen the methodological documentation, the GSBPM is supported on international standards such as SDMX, DDI and Dublin Core, which allow the documentation of metadata and indicators in the GSBPM model as well as the exchange of metadata with terminology and concepts common between different international organizations.

3. MAPPING THE DANE SPM WITH THE GSBPM AT THE PHASES AND SUB-PROCESSES LEVEL

Since the importance of the GSBPM was established in terms of strengthening of the national statistical production and comparability, the comparison was made between the DANE statistical production model (SPM), supported on the network of processes with the quality management system, and the GSBPM model in order to identify the modernization possibilities. In this way, we took into account the regulations, the network of processes and SPM organized by sub-processes at the level of the five phases.

With respect to the network of processes or value network, DANE recently performed some adjustments so as to see more clearly the role of lead entity and coordinator of the NSS, which DANE has, concatenating it more to the comprehensive production processes of the GSBPM:

---
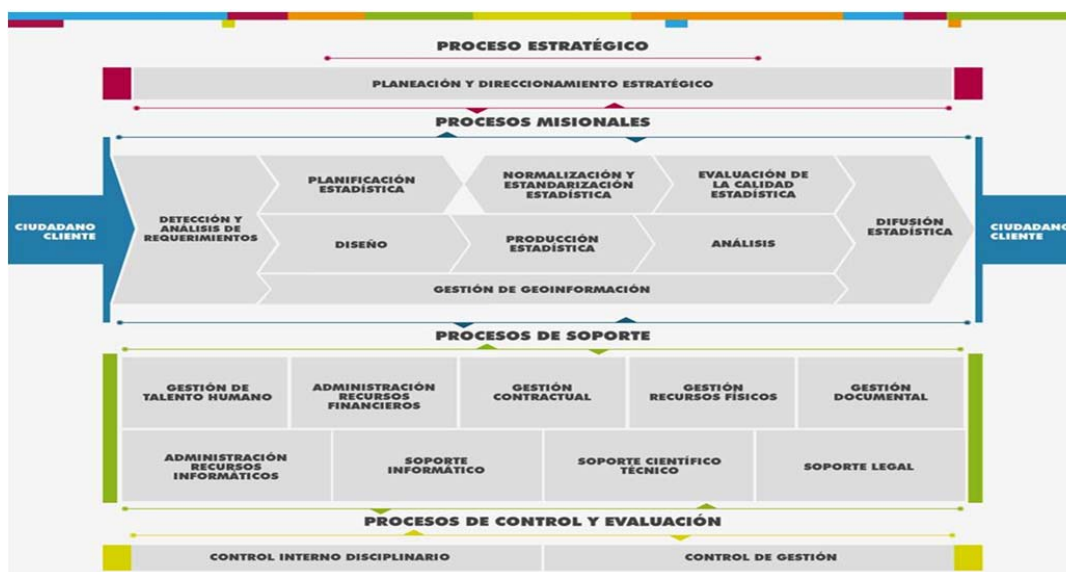
[3] Generic Statistical Information Model, 2013.

DANE value network. **Source**: DANE, 2016.

The current Statistical Production Model of DANE is presented below, which is referred to in the document as SPM:


DANE Statistical Production Model. **Source**: DANE.2016

This comparison was initially performed at the level of phases and was subsequently made at the sub-process level, examining the most significant differences and the impact in terms of comparability and interoperability. Firstly, a state-of-the-art SPM was developed, comparing it with the latest version of the GSBPM fostered by UNECE (version 5.0, December 2013). This mapping was done for each of the components of the levels:

• Level 1: 8 phases of the GSBPM production model version 5.0;
• Level 2: sub-processes of each phase.

This review allowed detecting in general terms, the making clear that a good design is not enough; it shall be accompanied by a proper implementation and a comprehensive evaluation process to verify whether the statistical program meets its objectives/potential for improvement in the efficiency of the existing network of processes, the review of the working structures that support the processes and the verification with respect to the management of the statistical quality requirements. Initial recommendations were generated by phases for the improvement of the DANE SPM.

The general results are presented below:
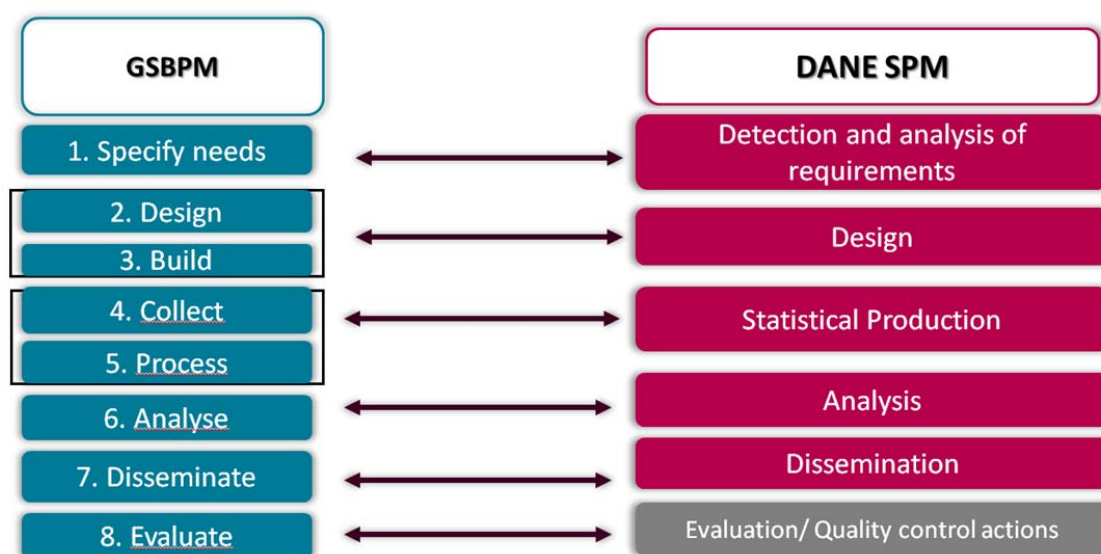
Relation of the DANE SPM with the GSBPM. Source: Author's own analysis

The mapping at the phases level shows that:

- The production phase in DANE SPM is equivalent to the collection and process phases in the GSBPM model.
- The design phase in DANE SPM is equivalent to the design and construction phases in the GSBPM model.
- Unlike the GSBPM model, the DANE SPM does not have the evaluation phase.
- There are components of the process that can be reused and integrated.

In turn, the sub-process mapping identified areas for potential improvement by specific phase and general for the DANE SPM, which were discussed with those responsible for the sub-process of all household statistical operations.

In order to perform this comparison, the project team formalized a work plan that consisted of four stages, which are explained below:

The first stage corresponded to the definition of social statistical operations (households set), which would be subject to the comparison proposed herein. For this effect, those responsible for each household statistical operation were established and the modernization project was socialized, led by DIRPEN. Also, some selection criteria were established, such as the potential for the integration of activities; the participation in the quality assurance process and the fulfillment of improvement plans; the capability to identify the frequency and coverage; use of dissemination standards worked in DANE (DDI, Dublin Core and SDMX) as well as recent redesign (so that the early phases of the GSBPM are taken into account).

As a result of this first stage the set of social statistical operations was established, which would be compared with the problems identified through the mapping of sub-processes of the GSBPM: National

Integrated Household Survey (GEIH), Quality of Life Survey, Cultural Consumption Survey, Political Culture Survey and Citizen Security Survey.

The second stage consisted of a review of the statistical operations selected, specifically relating to four fundamental aspects; (i) documentation[4] from the inventory of documents[5] published in the institutional management integrated system, as well as through the National Data Archive (ANDA); (ii), participation in the process of improving statistical quality; (iii) implementation of statistical standards; (iv) geospatial aspects. This second stage allowed a preliminary characterization of this set of statistical operations.

The third stage involved the construction of a questionnaire addressed to each of the teams of the selected statistical operations, i.e., thematic, logistics, samples, IT, and the ones responsible for the geostatistical, cartographic and directories component.

The fourth stage consisted of the application of this questionnaire: the consolidation, tabulation, and analysis of the information obtained. These actions allowed defining the traceability of the households statistical operations; specifying which are the activities performed in each of the phases and sub-processes; comparing the general findings that had been identified through the general mapping of the SPM and GSBPM, which in turn consist of specific challenges by sub-processes or work area; and identifying the opportunities for improving the statistical cycle.

3.1. Areas with possibilities for improving the DANE production model

Once this process has been performed, the main points for improvement were identified. These points include several sub-processes in the model and several phases, therefore some of them can be placed in what are called global processes in the GSBPM model. These identified points will be explained below.

3.1.1. Need for modernization of methods and tools for collection, processing and analysis

This point affects various sub-processes of the statistical cycle and is related to general aspects such as the possibility to analyze large volumes of information, promote the training of data scientists and even the methods used to carry out the data comparability between databases, leading to difficulties with respect to improving the analysis of information with more current tools and leverage other information and comparison sources. This point encompasses the following issues:

• Insufficient tools for the descriptive analysis and comparison of data for thematic experts.
• Updated data collection methods are required that reduce risks for the collected information.
• Difficulty in the use and development of applications for the Data Capture Hand-held Devices (DMC[6]).

---

[4] One of the objectives of this aspect was to identify the existing information, in order to avoid including related questions in the questionnaire and thus avoid the burden to respondents.
[5] Sampling designs, system applications, directories, methodological datasheets, forms and methodological documents among others
[6] For its acronym in Spanish

• The need for greater support on behalf of the IT division with respect to institutional initiatives for the training of officials.

3.1.2. Need for greater control of field operation costs

One of the main problems identified in the exercise with the thematic, logistics and sampling teams were the cost overruns in the field operation. This point goes from the most general (budget projections for the operations) to the most specific (related to training officials in the field or the use of electronic forms). This point encompasses the following issues:

• The human resource training plan generates displacement costs, which are quite high for the operation.
• Planning problems regarding the segmentation of the operation (there are every few re-counting staff at the regional offices).
• Cost overruns in field operation in terms of logistics and its budget projection. It is the main difficulty found with respect to the field operation. This, in turn, can be divided into several aspects:

- There are increased costs for re-visits: One of the main reasons for the re-visits is the failure in the training process of the interviewers and supervisors, resulting in incomplete forms or needs for verification of the information recorded.
- High operating costs in infrastructure, plant, expenditures in situ, utilities, etc.
- There is not a structural basis for measuring operational costs: In this respect it was noted that an update is required to the Tayrona[7] system, allowing a more efficient management of the financial processes related to the field operations.
- The data collection method most used is direct interviews, ..., which increase operating costs.
- There is no system that centralizes the costs of a field operation.

• Use of multiple applications for the development of a statistical operation, which generates higher costs and the burden to the respondents.

3.1.3. The framework for quality evaluation and improvement of processes is not clearly defined

This point indicates the lack of knowledge of the existing production model and the necessary distinction between this and the Quality Management System that encompasses much more than the model. Given the relative unfamiliarity with both, it is not clearly understood that the system allows evaluating not only the phases and sub-processes used in production but also the processes and sub-processes carried out in other areas that are "cross-cutting" throughout the statistical production: personnel management, financial management, knowledge management, etc. It is also clear in the presentation of the GSBPM model that it "may be used for the description and quality assurance of the processes based on surveys, censuses, administrative records and other statistical sources combined or not " (UNECE, 2016). This point approaches the difficulties that are mentioned below:
• The checkpoints with respect to the statistical production process need to be strengthened
• A greater monitoring needs to be performed with respect to the quality indicators relating to the field operation.

---

[7] Tayrona is a technological system, used by the Planning Advisory Office, which allows doing a better budget management in DANE.

3.1.4. Weaknesses in the standardization processes

Standardization is one of the general points covering more sub-problems. It is related to the weaknesses in the identification of information needs, the standardization of variables, surveys concepts and modules, as well as the management of metadata and microdata. In spite of the efforts made by DIRPEN, the standardization task is complex and very extensive, which requires increasing efforts in the case of adapting the GSBPM model. The specific points identified are listed below.

• A lack in the identification of the users' information needs (unsatisfied needs, disarticulation regarding the levels responsible for the detection of requirements and needs, underutilization of information on behalf of users).
• Inadequate implementation of dissemination standards of metadata and microdata.
• Weaknesses in the standardization and harmonization of concepts, variables and indicators.
• The operational scheme is not clearly documented and, therefore, there is not a well established procedure or institutional memory.
• The modules of the statistical operations pertaining to the same topic are not integrated.
• The microdata anonymisation process is not well articulated.
• Problems in the statistical critique.

3.1.5. Issues in the design phase

This point presents a big impact on the overall statistical operation, since the design phase includes all the sub-processes of technical and methodological design according to the requirements of the statistical information, even with respect to defining the statistical results and the reference framework that supports the operation as well as the collection and processing methodologies, and the tools needed for the fulfillment of the objectives; many of the difficulties in the development of the statistical cycle can respond to issues in the design phase. The specific points associated with the above are:

• The absence of unified criteria for the design of the technical bulletin.
• Issues in the design of the questionnaire.
• Weakness in the methodological design for the construction of the master sample of households.

3.1.6. Inconsistency between databases

The most notable inconsistencies between databases correspond to those generated by the IT and logistics work teams, which are reflected in disarticulated information flows and bottlenecks in the transmission of information that generate gaps and inconsistencies as well as a low integration of processes that are developed in each statistical operation. This point is related to the problem of the consolidation of the database that results from the collection process.

3.1.7. Update of frameworks

Problems have been identified in the geo-referencing of data resulting from the statistical operations and in the control of the recount. Problems were also recognized in the identification of dwellings and housings for the framework caused by their lack of updating. Over-coverage and under-coverage

problems were also identified partly because directories and statistical frameworks are outdated and there are delays in the development of the sample. In turn, the lack of a greater control over the recounting process is evident. Some thematic experts also noted that the statistical value chain or statistical cycle is not specified in the development of statistical frameworks.

• Problems in the preparation and recount activities for the update of the framework
• Difficulties with the update of directories and frameworks

2. Identification of opportunities

Through the workshops with the thematic experts of household statistical investigations in the fourth phase of the work plan, which is described in Section 3, approximately 100 opportunities were identified for the improvement of the SPM.

These opportunities were analyzed in a matrix[8] and the commonalities in them were identified. Given these structural similarities, as well as the areas for improvement outlined in the previous section, 11 general opportunities were found that are applicable to our project, and which can easily be placed in the GSBPM phases. Such opportunities are described below:

3.2.1. Human resource training

DANE staff is trained on topics such as human resources management and information technology management. These training processes are related to the collection, processing and analysis phases in the GSBPM model. This solution is intended to resolve the problem of overruns in field operation, to strengthen the different quality assessment processes and the lack of tools for descriptive analysis and comparison of data. In terms of the GAMSO standard, this opportunity is part of the corporate support level.

myrcompletion for questionnaires

The mapping and subsequent analysis of the questionnaire helped identify this problem not only in the framework of the collection phase within the GSBPM but allowed understanding it as a solution related to several problems: it can reduce operating costs since not as many interviewers would not be required per statistical operation.

In turn, the possibility of self-completion implies technological alternatives that do not involve physical forms (i.e., the development of online forms). The latter intends to tackle one of the specific problems that were identified: working with outdated information collection methods that jeopardize the collected information.

Some respondents noted that the self-completion process is a widespread collection method in other NSOs and which could be used in the entity. In order for the self-completion to be as effective as stated, the forms to be self-completed must be sufficiently clear so that the users do not have the option of making mistakes. This implies a very strong process with respect to the standardization of the structure of these forms, including the questions and modules of the questionnaire.

---

[8] The matrix and related information can be found as an attachment to this document.

3.2.3. Improvement of the Data Capture Hand-held Devices in hardware and software

Several years ago, DANE implemented the use of data capture hand-held devices (DMC) as a technological improvement for the data collection, processing and analysis processes. Both the mapping and the topical interviews conducted with experts allowed realizing that the DMC have been useful as a technological solution, but their potential has not been fully exploited. Some of the opportunities are related to their improvement, both in terms of software (e.g., to facilitate debugging) and the analysis of data consistency, the facilitating the development of platforms for recounting, as well as the improvements needed to transmit real-time data from field operations. This opportunity would require training interviewers with respect to the management of the DMC; it would improve the technological capabilities for analyzing, would avoid problems in the design phase since the development of questionnaires for DMC would facilitate data processing (including the analysis and consistency processes), as it happens with the recount. This opportunity would also allow updating the statistical framework in a more efficient way.

3.2.4. Design of an electronic form

One of the best-received technological alternatives among the surveyed teams was the development of an electronic questionnaire to be used in some of the household surveys. This opportunity represents a solution to many of the problems identified: it allows encouraging innovation in some phases of the GSBPM model; it allows for a reduction in costs since fewer interviewers would be required; and hard-copy questionnaires would be eliminated for the field operation. In order to create the electronic form it is required that response flows and the logic underlying modules and questions be included in their development, which implies them to be standardized. This standardization process will allow the development of unified modules for one or more surveys, which will improve efficiency in the development thereof.

3.2.5. Promote the use of technological tools

It was found that some processes could have considerable improvements if new technologies or technologies that can be improved are used: for instance, the collection and analysis processes undergo a considerable improvement by including the use of the DMC. Statistical dissemination would be improved by making use of new programs to display data by means of computer graphics, microsites, Java dynamic graphics, and pivot tables on the DANE website, among others. It was found that the use of these tools would have a direct impact on:

a) The decrease of cost in the field operations;
b) The update of statistical frameworks, especially in the update of collection instruments for the recounting of sampling units;
c) The review of possible inconsistencies between the databases handled by logistics and Information Technology Division (for household surveys);
d) The promotion of a statistical culture either through the website or in institutional scenarios;

e) The creation of automated bulletins[9] that facilitate the publishing of the periodic results of operations;

f) The creation of web-based statistical production manuals;

g) The promotion of the dissemination of microdata and metadata through the SDMX, DDI standards (among others), or through initiatives such as Open Data;

h) The use of technology to avoid the displacement of respondents when applicable[10].

3.2.6. Design of a single costs model

All the entities of the public sector need to make an efficient use of the scarce resources that are available. Specifically, DANE intends for the statistical operations to be carried out with the resources allocated without cost overruns due to a lack of planning. Respondents have noted that one of the central problems in DANE is the overrun with respect to the field operation. When field operations are developed, the training of coordinators and interviewers requires a great amount of logistical work and effort. One of the greatest difficulties is that the training plan involves moving trainers to regions where the field operation is carried out and interviewers are being contracted; also these trainings are repeated in each region where the operation is being conducted. These overruns include travel expenses for these trainers. It was also noted that the only collection methods being utilized are direct interviews with ideal or direct respondents, which increases the operating costs.

Furthermore, increases in costs have occurred due to the need for revisits; one of the main underlying reasons for these revisits is the problem involved with the training of interviewers, which in turn resulted in respondents delivering incomplete forms or the need to verify the information recorded.

High operating costs are also incurred with respect to infrastructure, plant, *in situ* utilities, etc., and there is not a structural basis for measuring operational costs. In this regard, it was noted that an update is required of the Tayrona system that allows a more efficient management of the financial processes related to field operations.

One of the most feasible solutions given was to centralize costs into a single model that ensures an effective control of the field operation budget and an adequate resources programming plan that facilitates online tracking as well as the integration between the Financial Reporting Integrated System and the Tayrona system in addition to meeting the requirements of *Colombia compra eficiente[11]*.

This model would also allow an update of the statistical frameworks, particularly for those municipalities that due to their size and location, require a very low turnover of the sample (as San Andrés Island).

---

[9] In DANE, a bulletin is a public document, related to each of DANE statistical studies, published following a strict schedule during the year, and in which aggregated data results of every investigation is shown.

[10] For some far places in Colombia, it is required that sources move to centralized meeting points (e.g., the nearest town with an internet connection) in order to complete the forms. In these places, a videoconference via Skype or other video streaming platform is recommended for conducting the survey.

[11] *Colombia compra eficiente* (Colombia buy efficiently) is the National Public Procurement Agency whose purpose is to develop and promote public policies and tools oriented to the organization and coordination of the entities involved in the public contracting and procurement processes in order to achieve a greater efficiency, transparency and optimization of the state resources. Taken from: http://sintesis.colombiacompra.gov.co/normatividad/documento/14208

3.2.7. Creation of modules and a bank of standardized questions by topic

The creation of modules and a bank of questions is one of the tasks of standardization that the mapping process shows as more feasible and simple to implement in household surveys, at least at an initial stage. This opportunity could eventually impact in a positive manner, in less time, the work of the teams with respect to the related topics. This opportunity is divided into two parts: a) A bank of questions; b) Survey Modules. Regarding the bank, it has been suggested that it be created with standardized questions that allow, in turn, establishing modules unified by topic.

Both tasks require a standardization process that is similar to that of concepts and variables, considering that the phrasing of the questions, i.e. their clarity is key to ensure that they can suit all audiences, whether or not they have the ability to answer them with the help of DANE staff. The integration of modules is a process which is relatively known and that has been implemented for several years in other NSOs and intends to integrate, in a first phase, some of the questions of statistical operations of a similar topic[12] e.g., (UNECE) or United Kingdom (Office for National Statistics, 2016).

This process requires hard work to map the modules and the questions that are capable of being unified in different topics, which allow facilitating the work of the operations involved. This problem is closely linked to the problem of correspondence between the questionnaire and output tables.

3.2.8. Strengthening of the SIMCO system

The Monitoring and Control Information System (SIMCO, for its acronyms in Spanish) was created to consolidate information in statistical operations and generate greater quality controls in order to have the necessary alarms[13] at the operational level (either municipal or departmental). Applied to the Third National Agriculture and Livestock Census, SIMCO has proven to be useful for monitoring the processes in it.

Several respondents believe that SIMCO is a valuable tool and that it could solve many of the problems to the extent that it allows a greater control over the sub-processes and phases of the model. Since the GSBPM presents a further breakdown in terms of these activities, the system is presented as an alternative to take control of these sub-processes following the GSBPM standard as a roadmap. Thus, the SIMCO would tackle the problems encountered in the design phase; it would allow strengthening the standardization solutions and developing a much more robust model for assessing the quality.

3.2.9. Establishment of improvements in the quality assessment process

One of the results of the mapping and discussion with the work teams is that tighter controls are required in the production processes in order to ensure the quality attributes for official statistics.

---

[12] A paradigmatic example of this integration is made with the harmonized concepts in the United Kingdom, where the questions and concepts of the initial modules have been unified by topic (e.g.(UNECE, undated))

[13] These alarms are in terms of the fulfillment of the operational plan objectives (simultaneously to the collection process, if an alarm arises in the municipality or department where the operation is being conducted, the corrective measures are taken)

The GSBPM mapping showed that traceability is possible and, therefore, it was identified as an opportunity to take the standard and develop a quality assessment system for the statistical production process. This would be coupled with the demands that are being made to ensure that the National Statistical System (NSS) has an evaluation system for the entities belonging to the NSS; its goal would be the quality certification of the official statistics they produce. This type of system would allow generating quality indicators for field operations in the context of the creation of standards and guidelines for the fieldwork.

3.2.10. Creating a harmonized variables system

DANE already has a harmonized system of concepts, but in various scenarios the need to standardize the variables that are used within the NSS has been stressed. The mapping allowed seeing that there are different variables used in a different ways for different purposes, which make the standardization process more complicated. Thus, a first task should be to unify a set of "nuclear variables"[14], for household surveys in a first stage, to see the impact of such standardization. This first action should bring two subsequent steps: firstly, the creation of a bank of standardized and harmonized variables and secondly the creation of new variables from those which have already been standardized.

This would be very important in order to measure the statistical potential of administrative records and alternative sources of statistical information such as the call data records (CDR) of mobile phones. In this sequence of tasks, standardization methodologies would be required to be applied in DANE in the first place and then in the other entities making up the NSS. These actions should allow establishing a harmonized variables system, which would be expressed in a query system opened to all users.

3.2.11. Standardization of a methodology for field operations

One of the major concerns of the work teams is the field operation plan, where some of the main points for improvement have been identified, starting with the need to standardize a methodology to carry them out. While the design and construction phases of the GSBPM explain all the sub-processes intended for field operation, the current DANE model does not specify them. The detail provided by the GSBPM allowed recognizing potential for improvement in the collection phase. For example, pilot tests were suggested for the recounting process in order to achieve a better coordination between the recounting and collection processes. This standardization provides for the review of existing operational manuals and the creation of those that do not yet exist, as identified above.

In order to reduce collection processes, the use of technological tools such as the DMC as already noted, was also suggested. Furthermore, an improvement in these two sub-processes would allow carrying out trainings so as to prevent the respondents from completing the forms in a wrong way (or even non-completion) in order to avoid non-response rates in the field operation. All these opportunities involve standardizing the sub-processes needed to carry them out and to make them

---

[14] For example, the nuclear variables providing demographic information: sex, age, country of birth, citizenship, and marital status. Those providing geographic information for localization: country, region, and address. Finally those providing socio-economic information: employment status, professional status, occupation, activity sector and highest educational attainment.

replicable not only in statistical operations in DANE but also those compiled by the entities making up the NSS.

The standardization of the field operation aims to reduce weaknesses in it and would allow more ordered updates of the sampling frameworks that are used by the NSS entities.

4. Some final thoughts about the exercise

The comparison between the GSBPM model and the DANE production model allowed finding similar phases and sub-processes, at least partly due to the fact that DANE has closely followed the developments conducted by Statistics Canada in this regard.

Furthermore, some of these sub-processes (e.g. some sub-processes in the processing phase, such as review and validation or in the collection phase such as preparing the collection and carry it out), and some phases (e.g., design phase) require a revision at the thematic, logistics and IT levels. It was also possible to see that many of these challenges are cross-cutting and go beyond the phases that the model proposes; they can be traced to the support processes that in the GSBPM are known as overarching processes. For example, the inconsistency between Logistics and IT databases can be traced, as a problem, from the design and construction phases. Even though the problem of overruns can be located in individual sub-processes, they also can be located within the crosscutting processes and ultimately they affect the entire statistical operation.

The mapping facilitated the identification of challenges in the updating of sample frameworks and in the development of statistical standards that can be tackled very quickly whereas some may require more development (for example, the implementation of technological solutions for the improvement of sub-processes in the collection phase, which would require standards previously generated by DIRPEN).

The mapping also allowed identifying opportunities for improvement in several areas, for example, the standardization of the questionnaire modules in household operations, as well as a single methodology for field operations. It also justified that the development of a harmonized variables system should follow steps similar to those that have been covered with the harmonized concepts system since the sub-processes that it affects are almost the same.

In summary, the solutions to the challenges that the entity faces can be divided into three main areas: human resource training, enhancement of technological tools and the strengthening of the design and implementation with respect to the statistical operations involved. The solutions of these three fields, coordinated, would allow the sub-processes involved to undergo considerable improvements in both efficiency and effectiveness, which should be reflected in the scope of what has been called the industrialization process of the statistical production.

It is clear then that DANE statistical operations can be improved in the sense that they can be industrialized, basing this improvement in a standard such as the GSBPM model. Nevertheless, the improvements that the model allow do not necessarily imply their adoption or even their adaptation, because the GSBPM model is a general good practice of the statistical production and as such can be subject to adaptations or rejection when it ceases to bring the positive effects seen with its adoption so far.

As the modernization project has been designed, the mere identification of opportunities is the first step. As a further task, the team has established criteria for prioritizing such opportunities and defining strategies for the consolidation of DANE modernization plan.

BIBLIOGRAPHY

Community, S. (2016). *SDMX.* Retrieved 2016, from https://sdmx.org/

ESCAP. (2016). *Economic and Social Commission for Asia and the Pacific.* Retrieved 2016, from http://www.unescap.org/our-work/statistics/modernization-statistical-production-services

EUROSTAT. (2009). ESS Standard for Quality Reports. Methodologies and Working

papers.

Javorsek, M. (s.f.). *Modernization of Statistical Production and Services.* Retrieved 2016, from http://www.unescap.org/sites/default/files/Modernization%20of%20producton%20and%20services.pdf

Mundial, B. (2016). *DDI.* Retrieved 2016, from http://www.ddialliance.org/

Office for National Statistics. (2016). Harmonised Concepts and Questions for Social Data Sources. *Harmonisation Standards*. London. Retrieved Junio 15, 2016, from https://www.ons.gov.uk/file?uri=/methodology/classificationsandstandards/harmonisationstandards/introhstandv4nov11_tcm77-247522.pdf

Olenski, J. (2003). Global Standard for harmonization of Social Statistics, with special

reference to transition and globalization processes. Expert Group Meeting on Setting the Scope of Social Statistics. United Nations Statistics Division in collaboration with the Siena Group on Social Statistics. New York. SSDIS ESA/STAT/ AC.88/10 de abril de 2003.

Ortega Venzor Alberto. (2009). Los estándares estadísticos internacionales y la

armonización de las estadísticas nacionales. (*The international statistical standards and the harmonization of national statistics*) INEGI Agosto 2009 Director General de Vinculación Estratégica, INEGI, México- SDMX 2008.

UNECE, (2013). "Towards a Generic Statistical Business Process Model". Steven Vale.

UNECE. (2016). *CSPA v1.5.* Retrieved 2016, from http://www1.unece.org/stat/platform/display/CSPA/CSPA+v1.5

UNECE. (2016). *GAMSO V1.0.* Retrieved 2016, from http://www1.unece.org/stat/platform/display/GAMSO/GAMSO+v1.0

UNECE. (2016). *GSBPM V5.0.* Retrieved 2016, from
http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model

UNECE. (2016). *GSIM.* Retrieved 2016, from
http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model

UNECE. (2016, May). *UNECE.* Retrieved from GSBPM v 5.0:
http://www1.unece.org/stat/platform/display/GSBPM/II._The+Model

UNECE. (s.f.). *Primary set of harmonised concepts and questions.* Retrieved abril 2016, from UK
ONS:
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/harmonisation/primary-set-of-harmonised-concepts-and-questions/index.html

UNITED NATIONS. (2006). Common metadata framework. Part A. United Nations Commision for
Europe. July, 2006.

UNITED NATIONS. (2009). National Institute of Statistics and Geography of México: Global
statistical standards and harmonization of national statistics. Statistical Commission. December,
2009.

UNITED NATIONS. (2010). Report of the Committee for the Coordination of Statistical Activities on
the Inventory of Global Statistical Standards. Statistical Commission. December, 2010.