# Implementation of General Statistical Business Process Model in the administrative source

Elsa Dhuli, Institute of Statistics, edhuli@instat.gov.al

**Abstract**

*The quality of production of official statistics is a contemporary demand which is growing rapidly by all countries, no matter what their inputs, statistical surveys or administrative sources are. Increasing the use of administrative resources as a primary or substitute element first requires first-rate quality assessment. Given the fact that the common approach is the measurement of the input quality and the statistical product process, when the main source is the administrative source, this approach takes on other dimensions. Generally speaking, the quality of each processing phase up to the realization of the statistical product has its importance, but the quality of the collection phase of administrative information and the implementation of statistical models in their holders also significantly guarantees the quality of other phases of the process.*

*This paper aims to outline the implementation of General Statistical Business Process Model in the administrative source system at some of the phases to guaranty their quality to be usable as primary source for producing official statistics. The analysis of the quality of each phase of the statistical process from the administrative source of the experience of developed countries and the recommendations given for further improvement is a positive approach for enlargement countries.*

**Keywords:** Quality, administrative source, General Statistical Business Process Model (GSBPM)

# 1. Introduction

The production of official statistics from administrative sources, not only from statistical surveys, in Albania, is a new approach. This is in the framework of having higher benefits, better coverage and lower production costs for official statistics. There is an increasingly need to have alternative sources for producing monthly statistics related to turnover, wages and salaries, number of employees, etc, or having a combination of sources such as statistical surveys and administrative sources.

The reasons we use administrative sources are three: i) as population frame for statistical surveys, with a view for sampling design and creating and updating statistical registers; ii) aid for imputation or non-response; iii) as full or partial substitutes for statistical surveys.

By itself, administrative sources are organizational units, and such sites, may be public or private. Administrative sources, where owners of which may be the state or private, are not created for statistical purposes. Regardless of ownership, administrative sources have been built to follow administrative processes and procedures. Meanwhile, they are also a potential source for statistics production. "[UNECE, 1998].

Some institutions in Albania that are potential sources for producing official statistics are:

* Tax Authority (Value added tax)
* Social Security (social contributions)
* Custom Authority (import/export)
* Civil Status Office (civil status register)
* Health Insurance State (health insurance)
* Road Services Authority (inventory of road vehicles)

The above-mentioned organizational units produce indicators for measuring the current situation in Albania for different area, as well as measuring their work performance.These sources are not used for producing statistical indicators, and even more qualitative indicators for official statistics. As such, the documentation of their information is made according to the rules and standards they have created for or adapted to those in developed countries.

As we pointed out above “an administrative data source can be used for other statistical purposes than being the primary source of data for statistics” (Eurostat, 2003), for this reason, the quality requirements and standards for an administrative register will depend on the role of the register in producing official statistics. Based on the purpose of using these sources for surveys or direct production of statistics they are dependendent also from the definition of dimensions of quality. So, if they are to be used for surveys, they need to have good coverage, ability to identify population units, contact information, domain indicators and information on auxiliary variables. But, if they are used to produce statistics, the preconditions of use of administrative sources such as coverage, the ability to identify population units, key value of major variables, domain indicators are the most useful.

The production of official statistics passes through business processes. The GSBPM is a standard framework that is used to describe and define the set of business processes needed to produce official statistics. As it provides harmonised terminology the GSBPM has help National Statistical Office to modernise statistical production processes, as well as to share methods and components with other statistical agencies. The GSBPM is used for integrating data and metadata standards and as a template for process documentation. Also, it is use for harmonizing statistical computing infrastructures, and to provide a framework for process quality assessment and improvement. [UNECE, Version 5.0, December 2013, p.3].

The ES Code of Practice addresses the main quality issues related to statistics based on administrative data, especially relating to accuracy and comparability principles. ([A.Baigorri; R.Laux;W.Radermacher, Eurostat 2009). But not only, in the latest study related to quality indicators used in the Generic Statistical Business Process Model, quality indicators produced by statistical surveys or administrative sources are both qualitative and quantitative. [UNECE, v.2.0, October, 2017,].

# 2. Implementation of GSBPM for increasing the quality of produces and uses of statistical products

### 2.1 General issues

One of the main questions of Institute of Statistics addressing to administrative sources, is: If the documentation of proceses is done from administrative source according to the GSBPM framework, would it bring a qualitative data?

Quality is a matter not only of organizational units, but also of the processes they follow in the production of products as well as the products themselves. Production of quality indicators to avoid possible mistakes during data processing processes would eventually bring about quality products. The reference framework for documentation of the statistical processes in phases and sub-proceses on which this paper is based is the Generic Statistical Busines Process Model (GSBPM), version 5 [UNECE, 2013].

The purpose of this paper is to demonstrate how this framework is implemented in INSTAT. But during its implementation, where administrative sources are used in many phases, for the purposes mentioned above, this paper tried to explain how this framework could be implemented from administrative sources. Can different organizational units, in order to have as much official qualitative statistics as possible in the everyday life of any of us, use the documentation of their processes, regardless of the purpose of the creation, according to the standards of the GSBPM framework?

Walgren and Walgren, 2007 in their study have taken the example of the three main processes that pass an industrial product (whatever is it), in parallel, these processes pass a statistical product. There are two actors who are interested in the quality of each stage, the producer and the users. With the same parallelism, the manufacturer wants to produce a quality product that has a significant market place; the customer wants to buy a quality product. In this context, the quality of production of the statistical product is important both to the producer of statistics in the first two phases (inputs and processing) and to their users (outputs). The simple analogue to the production of industrial products, illustrates that the quality of administrative data has to be viewed from two different views, from the viewpoint of the user of statistics and from the view of the producer of data for statistics. [Walgren.A and Walgren.B, 2007].

### 2.2 National experience in implementation of GSBPM

By the Article 4 and 5,, of new statistical law, No 17/2018, “In order to ensure the official statistics quality and their reliability in the public eye,” and “To ensure the quality of production, by following these quality principles: “Relevance”, “Accuracy”, “Timeliness”, “Punctuality”, “Accessibility” and “Clarity”, “Comparability”, “Coherence”, “Affiliation” “Transparency”, the official statistics procedures are guided by the principles of the European Statistics Code of Practice and shall be developed, produced and disseminated guided by unified standards and harmonized methodologies. INSTAT has started the documentation of statistical processes with the Generic Statistical Business Process Model V.5. GSBPM is used to have a standard framework in all statistical process and products produce from INSTAT, for the description and definition of the processes that are required to be used in production of official statistics with harmonized terminology, for computing the same infrastructure. Documentation of processes production is a basic element for any statistical office, but it also integrates metadata data in a standard way. This framework is applied for statistical products produced by surveys and administrative sources.

One of the indicators that highlight the dimension of quality, relevance, in Albanian case, is the existence of a unique identification number for enterprises in all administrative institutions related to businesses All databases are centralize in one common government security platform. The delivering of those to interested party is doing under strict confidentiality rules. While, distribution of statistics to the user is online and every user (in this case the administrative source) in the cost-effective process documentation, provides cost-free statistics (exemptions make special requests), while the documentation by the statistics requires cost estimation of the transmission system.

### 2.3 The importance of documentation of procceses according to GSBPM from administrative sources

Meanwhile administrative sources themselves, document administrative processes and have built flowchart for their purposes, according to a standard set by themselves. The same approach of administrative sources to statistics would facilitate their study and analysis.

The need for qualitative statistics is a contemporary demand that takes special importance. All actors in the country cooperate to increase the quality of their production, at institutional level, production process, and final product. If the statistics are users of administrative source data, the latter are user of statistics. As an administrative source, statistics are the users of the information that each of them carries. Statistics and administrative sources have one approach: Administrative sources output serve as input for providing statistics; on the other hand, statistical output can serve as input for Administrative sources to draft their policies. The intermediate stage of processing is used for individual purposes of each of them, but in common interest for both actors.

The GSBPM system reflects the phases and sub-phases where a statistical product passes, regardless of the statistical or administrative survey source. This model is used internally in INSTAT to set standards in the processes through which a statistical product passes. Description of process-oriented, general-purpose production, metadata data and metadata integration, harmonization of the statistical structure and calculation, with the sole purpose of providing a structure for the process of quality assessment and improvement there in where gaps are noticeable. Quality dimensions and relevant indicators are clearly defined for both sources, [UNECE, October, 2017], for each dimension. In the ”Data Quality Assessment Tool for Administrative Data”, there are 6 quality dimensions mentioned for assessing administrative sources: Relevance (timeliness), Accessibility (comparability), Coherence, Interpretability, Accuracy, Institutional Environment. [W.Iwig et al., 2013].

The phases and sub-phases of the GSBPM, which logically have to follow the linear line, in special circumstances, aim to transcend them from one sub-phase to another, especially this is where the source of information is administrative. (UNECE, 2009).

Standards over which the administrative processes pass are not the same as the standards that pass the statistical processes but the coordination, combination and effectiveness of their use in phases or sub-phases where they interrupt each other requires equal attention and perspective from both sides from the sole purpose of producing quality statistics, both for the producer and the user.

### 2.4 Phases / sub phases by administrative sources

In the first phase, *Specify needs*, the necessity of using the administrative source is noted in the sub-processes: Identify needs, Confirm needs, Identify concepts, Check Data Availability.

One of the purposes of the phase *is* the assessment of the dimension of quality, Relevance, for statistics derived from surveys and administrative data sources, two indicators should be defined: To what extent have stakeholders been identified and included in discussions about statistical needs? and To what extent has relevant supporting documentation been collected? (UNECE, October 2017). If the first indicator is relevant to administrative sources, the second indicator is more relevant to statistics.

Documenting this phase from the administrative source would determine the need for new statistics or details of produced statistics. Meanwhile, statistical documentation has been analyzed and the need for administrative sources is evident.

*For example*, administrative sources as Tax authorities need statistics on Sales indices at the 4-digit level according to NACE Rev. 2, statistics need detailed business sales datasets to enable production of indexes of sales according to the required detail.

*Consult and confirm needs,* would be documented by Administrative sources, for new statistics or additions to existing ones, would be of a qualitative value for statistics if their needs would be confirmed in a detailed level understandable for statistics, time and reason. [UNECE, October, 2017].

*Identify concepts*, is the most important process to be documented by the administrative source, first because it will describe in detail the variables, part of its data system, eg net sales, by categories (large business, small business,VIP), number of employees, according to the description (paid employment, self-employed), payments made (according to established thresholds). The description of concepts and definitions is standardized, but the combination of both descriptions, following the standard way (GSBPM), increases the level of clarity of exactly what the variables describe where they merge and divide.

*Check data availability*, by the administrative source would document the level of availability of statistics for their analysis.

At the moment of its creation, the legal framework of the organizational units has been designed to regulate the administrative operation of these resources, the obligations and the rights of the involved actors, and it has not been established to determine the obligations towards the statistics. The most important quality dimensions are the statistical confidentiality and cost effectivnes to analyse and take into consideration.To understand the purpose of the obligation of any administrative product it is important to study the by-laws that are drafted for their implementation and in support of these laws as well as the materials related to them and published by them, alternatively (Laitila, Walgren and Walgren, 2011). INSTAT for strengthenin the cooperation with owners of administrative data and regulate relation for data exchange has signed during 2017th, 19 Memorandum of understanding.

Obtaining detailed information from the administrative source, in this case the detailed sales at enterprise level according to economic activities performed by the enterprise (in the case of performing two or more activities), requires the need the completion of the declaration form from the businesess at the activity level rather than the level of revenue / total sales, identifying only the main activity, and not the other secondary activities. This complicates the statistics work for statistics for calculating short-term indices from the administrative source using the kind activity unit as it only allows its calculation at the enterprise / legal unit level.

Coordination of development strategies in terms of improving the quality of statistics by both actors could also be part of co-operation agreements.

*Design output,* in the documentation by the administrative source as an information source, defines their administrative purpose, concepts, collection tools, and the operational processes they carry out for their purposes, facilitating the work of statistics on meaning fair and accurate output of them. The length of time series of data is the important quality dimensions needed to document by owner of data.

Once, "the metadata of Administrative sources describes, or points to documents that describe, among other things, the administrative procedures that create the data, any important administrative events relevant to the data and definitions of concepts, variables and the population they refer to”. [Froeschl and Grossmann,1999]. As producers, documentation of variable descriptions, classification used, etc in this phase from the administrative source would require to manage metadata related to them considering as important quality dimension. As user, each statistical product had the relevant metadata as well as the quality report with the respective indicators. INSTAT has published quality report for 19 statistical products.

The workflow of the sub processes from data collected to dissemination of out from owners of the data for statistical reason giving an overview of the whole process will increase the efficensy and effectiveness of both actors. The soundness of implementation, accuracy and reliability, accessibility and clarity are the most needed quality dimensions to take into considerations.

In the *Process and Analyse phase*, it is important to receive the documentation by administrative sources, Classify & Code, according to GSBPM. For example, the documentation of economic activity part of their system according to NACE Rev.2, the way they have implemented the NACE classification, the coding quality and the code revision by them, reduces the work of statistics, in case of using administrative sources for the statistical purpose, for the accuracy of the codes. In frame of increasing the awareness for having better economic code from administrative data, INSTAT has organised training with tax authority and National business center, in center and local level, (around 200 persons).The combination of documentation of this subject from Administrative sources and Statistics is indispensable. The Compliance Rate Indicator, for the sound quality dimension of the methodology, calculated by both actors is neccesearly.This is to increase the transparency of statistical products.

In *Analyse phase* the documentation by organizational unit of the administrative data would include the description of the quality control methods it has performed to ensure that its data are accurate. Calculation of the report on percentage of Errors in the dataset would be an important statistic evaluation.

The *Output*, whose documentation by the administrative source, implies the description of administrative data quality characteristics in the preparation of the final fileused as Input for producing official statistics.

*Validate outputs*, meant by owners of administrative data, documenting components such as: Clarity; Administrative concepts; Coverage; Reference time; Data freshness; Errors in the data; completeness; Record matching ability; Confidentiality and privacy protection; Compatibility between file formats; Comparability of administrative data in time; Envisaged uses of the data. The evaluation of these elements is a guarantee for statistics on the use of administrative data for statistical purposes, especially by relaunching one of its main uses, amongst the three, use as the primary source. To confirm that the quality level of the dataset prepared to transmit to statistics is fit for the purpose they will use the collecting supporting information, technical notes and other internal information is necceseary. (Tab.1).

**Tab.1 Quality dimensions for each phase**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GSBPM phases** | **Quality dimensions** | | | | | | | | | |
| **Relevance** | **Accuracy and reliability** | **Confidentiality** | **Cost effectiveness** | **Coherence and comparability** | **Managing metadata** | **Soundness of implementation** | **Timelines punctuality** | **Accessibility and clarity** | **Methodological soundness** |
| Specify needs | x |  | x | x |  |  |  |  |  |  |
| Design |  |  |  |  | x | x | x |  |  |  |
| Collect |  |  |  |  |  |  |  | x | x |  |
| Process |  |  |  |  |  |  |  |  |  | x |
| Analyse |  | x |  |  |  |  | x | x |  |  |

The evaluation of processes followed by owners of the administrative data for statistical purpose is needed. This is a future task for INSTAT.

# 3. Conclusions

In this paper are highlighted the phases and sub-phases that are necessary to be documented from owners of administrative data with GSBPM. While this framework is used by statistics for documenting statistical processes, regardless of the primary source of data, statistical survey or administrative source, it would be very important to use it from administrative sources.

Documentation of administrative data processes according to GSBPM has its advantages and disadvntages. The advantages of documenting the administrative process according to GSBPM framework will increase the quality of data from sources, will have harmonize concepts and definitions, the same framework for comparing the data, the same methodology used, accuracy data, etc. The disadvantages are that the implementation of this framework could bring new legal actions or improved existing one, increase cost from organisational unit, but at the end will have qualitative statistics needed for policy decisions for better life.

The preparation of internal quality reports by administrative sources and the calculation of quality indicators would increase the usability of administrative sources not only for administrative purposes. Realization of documentation according to GSBPM, by administrative sources, requires supports from statistics.

Support that is among others accomplished with the adoption of legal acts, if necessary, the signing of memorandum of understanding, training undertaken by statistics what is and how to use GSBPM (documentation and metadata), institutions serves as a tool to identify potential source through data they gather for their needs or services they offered, training on classification and codification, how to draft internal quality reports.

# References

K.A.Froeschl and W.Grossmann, (1999), Metadata as a Tool to Improve Quality of the Statistical Production Process.

Eurostat (2003), Quality assessment of administrative data for statistical purposes, Working Meeting, Doc. Eurostat/A4/Quality/03/item6

A.Walgren, and B.Walgren, (2007), Register based statistics; Administrative data for statistical purposes.

Eurostat 92009), ESS, Standarad for quality reports.

A. Baigorri; R.Laux: W.Raderm acher (2009), Building confidence in the use of administrative data for statistical purposes.

T.Laitila, A.Wallgren and B.Walgren (2011), Quality assessment of administrative data, research and development strategy.

W.Kloek and S.Vaju (2013) The use of administrative data in integrated statistics”,NTTS, 2013, Brussel, Belgium.

UNECE (2013), Generic Statistical Business Process Model, V.5, December, 2013.] (<https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0> ).

W.Iwig et al, (2013),Data Quality Assessment Tool for Administrative Data.

M.Agafitei, et al., (2015), Measuring output qualityfor multisource statistics in official statistics: Some directions”, Eurostat.

Eurostat, (2017), Revised European Statistics Code of Practice, November 2017.

UNECE, v.2.0, (2017) Quality indicators for the generic statistiacal business process model derived from Surveys and administrative data sources, With contribuitor A.Born et al., ,pp 5- 56.

INSTAT (2018) New Law on statistics, No18, date 05.04.2018, Available at: http://www.instat.gov.al/en/about-us/statistical-law.