Introducing Enterprise Architecture Framework in Statistics Poland

J. DYGASZEWICZ*, B. SZAFAŃSKI**

j.dygaszewicz@stat.gov.pl
boleslaw.szafranski@wat.edu.pl

*Central Statistical Office of Poland
Al. Niepodległości 208, 00-925 Warsaw, Poland

**Military University of Technology, Faculty of Cybernetics
Kaliskiego Str. 2, 00-908 Warsaw, Poland

Article is devoted to the modernization of the statistical production process. The starting point for formulating the principles and methods of the modernization is to establish a framework architecture. An in-depth analysis of all relevant aspects arising from the holistic approach to the production of statistics is essential to extract the key business issues. Business needs are the foundation for the formulation of a coherent and transparent guidelines, demands, legal and technical requirements, both domestic and international which are the basis for the construction of the enterprise architecture framework and the development of solutions used to carry out the tasks of public statistics. The process of constructing architectural framework requires the use of models and methods used in scientific research, in particular arising from the achievements of enterprise architecture and modeling tools for object-oriented computing.

Keywords: enterprise architecture framework, statistics survey, GSBPM, TOGAF.

1. Introduction

The mission of public statistics is to deliver reliable, independent and high-quality statistical information about the phenomena of society, economy and natural environment, thus responding to the needs of domestic and international recipients.

At present, in the age of IT innovation, the reality creates new challenges for public statistics. In order to implement them and strengthen the challenges and the growing needs of recipients, changes in the way the public statistics operates are necessary. It is, after all, not only the key element functioning within the IT system of the state but also European statistics. What is also relevant, is the exploration and appropriate use of new possibilities that are, at present, offered by the growing open data sources and dynamic development of new technologies.

The most relevant challenges facing public statistics include:
1. Customers’ raising needs for good quality data and also proper metadata;
2. The needs of cost effective data collections and reducing respondents’ burden;
3. Expectation time shortage of the statistical production process;
4. Orientation on cooperation and communication:
   • external (with the statistical stakeholders)
   • internal (sharing the knowledge and methods among the statisticians);
5. Surrounding changes (ICT, organizational, legal) which would face the challenges.

Concept of statistical surveys organization in the Central Statistical Office (CSO) of Poland requires taking action aimed at increasing the efficiency of the statistical production process and accompanying organizational and coordination processes. To achieve this goal the Enterprise Architecture Framework has to be introduced and lot of activities must be taken, out of which three main ones, taken with the view of making the organization more efficient, are as follows:

1. Identifying and establishing processes within the particular stages of official statistics, based on Integrated Statistical Production Process (ISPP), supported by Generic Statistical Business Process Model (GSBPM) [1], appropriate to reality and the needs of Polish official statistics;
2. Specifying the vision of implementing statistical survey as standard of integrated statistical production process taking into consideration metadata objects;
3. Standardizing organizational processes, which is based not only on ensuring uniform solutions regarding data collection, processing and dissemination but also on implementing the entire statistical process according to uniform principles and procedures;
4. Preparing the procedures that make it possible to automate planning processes and survey implementation;
5. Implementing the concepts, rules and procedures in stages, with regard to the possibilities of human and financial resources;
6. Defining goals and requirements for the metadata system supporting management of integrated statistical production processes;
7. Implementing the meta-data concept that enables the use and creation of meta-information at every stage of the statistical process as well as management of access, users and the quality of meta information.

The Concept of the introduction enterprise architecture framework in the CSO of Poland covers the entire transformation programme from the current surveys organization model to the on Integrated Statistical Production Process supporting by the Generic Statistical Business Process Model [1].

2. Enterprise Architecture Framework

Enterprise architecture is a formal description of the organization components structure and function, the relationships between these components as well as the principles and recommendation for their overtime creation and development. The organization component is any component used to construct the organization (it can be people, processes, physical structure as well as IT systems).

Typically, enterprise architecture framework consists of 5 main components (ref. TOGAF [2] – The Open Group Architecture Framework) such as:
1. **Enterprise Architecture Principles** – the collection of permanent principles based on the organization development strategy, which shows the comprehensive needs of the organization in creating IT solutions, consistent with the Architecture Development Method (ADM);
2. **Business Architecture** – defines the business strategy, ways of managing the organization governance, organization structure and key business processes as well as relationship between the components;
3. **Data Architecture** – describes types and data sources necessary for the organization functioning;
4. **Application Architecture** – describes individual applications to be deployed, their interactions, and their relationships to the core business processes of the organization;
5. **Technology Architecture** – describes the technical infrastructure which is the basis for functioning of key software (this includes operating systems, database management systems, application servers, hardware and communications infrastructure, etc.).

Organization of public statistics should create a coherent system, i.e. a set of elements which are mutually correlated and which construction, functions and links ensure the most efficient functioning of the entire structure [2]. Efficient functioning of statistical survey organization means in particular delivering the expected products and statistical services to the recipients, while observing the required number, quality, timeframe and budget. Modern public statistics imposes a strong pressure on increasing that efficiency, i.e. it regards the need to deliver new products faster and cheaper with less burden on the respondents.

In such a context public statistics can be equated with “the enterprise” (Each statistical organization is a factory of statistical information) [3], in which statistical surveys can be acknowledged as the so-called basic operations of the enterprise – having the decisive influence on its efficient functioning.

3. General Model of Public Statistics

As mentioned earlier, the organization of public statistic surveys should is a coherent system creating high level own enterprise architecture. Among the key elements of this system one should mention:
1. Work process;
2. Organizational structure;
3. Tool infrastructure (=> IT);
4. Knowledge base;
5. Recipients of products and services of public statistics;
6. Suppliers of services and data for public statistics;
7. Legal regulations.

The entire system requires efficient coordination and protection of sources of financing.
The above mentioned elements remain related. Analysing the statistical surveys system organization and in particular potential possibilities of improving IT components, it seems that it is not possible to obtain significant positive improvements concentrating only on a single element of the entire system. This is the result of the elaborateness of the public statistics organization system, the elements of which influence one another and every change of each element leads to a change in the entire system’s functioning.

Potential changes in such enterprise architecture, especially those regarding the improvement of efficiency, should be prepared comprehensively with regard to each of the above elements.

4. The current model of statistical surveys organization systems

The current model of organisation of surveys in the Polish statistics is continuation of tradition and experiences of the previous system of public statistics. It is of character of the model known as "stovepipe", so statistical surveys are carried out by field and are within authority of various original organisational units (approx. 20) – departments and specific statistical offices – from the beginning to the end of a statistical production process, namely from planning and designing the survey to making available results of the survey. The leader unit specialising in a given thematic field plans, carries out surveys and draws up results of it in accordance with prepared procedures, separate for each survey.

In so functioning organisation of the statistical process, eight main stages of data processing can be identified: specification of needs, designing, building, data collection, data processing, data analysis, data dissemination and ex-post evaluation. The above-mentioned main stages of particular surveys are basically in accordance with stages determined in the so-called generic statistical business process model (GSBPM) – showing a standard approach inter alia towards description of a process of the statistical production. However, in the realities of the Polish public statistics realisation of this model occurs separately with regard to every survey and is executed in an individual manner, as a result of own implementation of a process of the statistical production, e.g. often with significant overlapping the boundaries between particular stages of realisation of the statistical survey.

With such organisation of the public statistics surveys, problems occur which in large measure are a result of the so-called silo effect. There is some kind of limiting – as part of separate organisational units - only to own tasks, attempting to implement them – in own individual manner supported by dedicated IT applications – as well as possible, but using "common" methods of work. The result of such a situation is that as part of an organisation of the public statistics surveys, the following problems can be, among others, identified:

1. High diversity (including lack of “compatibility” and standardisation) of work processes and IT tools;
2. Redundancy and often lack of specialisation in executed activities, in particular redundancy concerning the scope of information obtained from respondents;
3. Duplication of competences being a result of, for example, lack of specialisation of activities;
4. In connection with every survey carried out, a separate IT system is developed and maintained;
5. Excessively complex IT environment, difficult to maintain and develop, consisting of about 200 dedicated and independent IT systems, created for the needs of particular publications. The result of such an organization of IT environment are additionally:
   - significant redundancy/duplicating of functionality
   - increased expenditure (finance, and staff) on maintenance and development of the IT environment
6. Collecting data independently for particular surveys and the impossibility to share them for the needs of other surveys (data gathered in several separated bases);
7. Predominance of surveys for which data are collected via forms and questionnaires directly from the respondent with insignificant use of information available
for instance in the public administration registers;
8. Lack of a comprehensive metadata service.

The above-described problems are visually shown in the figure below (Figure 2).

![Fig. 2. Stove-pipe model of current survey organization](image)

Taking into consideration the above deliberations, it can be said that as part of such organisation of the public statistics surveys, there are a two hundred, mostly independent, so-called production lines – dedicated for each statistical survey.

In this situation assessment of the organisation of surveys in the light of the generic model was clear and induced to make a decision to introduce an architecture framework establishing integrated model of organisation of the surveys, with the use of rules of the generic model (GSBPM). Currently it is a desired direction of organisational changes of the public statistics in almost all National Statistical Institutes in the world.

5. Framework of the corporation architecture

A holistic view on the organization of statistical data processing and statistical corporation provides methodology defined as Enterprise Architecture. To represent a systematic mode of action of enterprise architecture action the Enterprise Architecture Frameworks are used, which are the specific set of models, methods, instructions and specifications used for the construction of enterprise architecture in organizations.

The framework of corporation architecture in the CSO of Poland covers the entire transformation from the current surveys organization model, based on a stovepipe model, to the Integrated Statistical Production Process. It requires adequate preparation of the institution, and in the first place carrying out an analysis of previous work processes particularly for identification of “similar” activities with “analogical” results and which required “similar” competences. Based on this – in particular based on GSBPM – a uniform and coherent project of standardised processes corresponding to present production lines of the statistical surveys should be prepared. Introduction of a new model of organisation of the statistical surveys, aiming to optimise, re-design and standardise work processes, data architecture and application architecture including IT infrastructure.

Change of the model of organisation of the statistical surveys includes many aspects of functioning of the statistics – it thus requires many activities and preparation such as:
1. Defining production processes in specific research realities;
2. Integrating surveys in different domains of statistics – it is difficult to in process carry out 250 topics of surveys at present realised in the system of Polish statistics;
3. Building a system of metadata as the basis for integration and standardisation;
4. Preparing a concept and implementing a new IT integrated infrastructure for drawing up surveys;
5. Adapting institutions to a change of the organisational structure adequate to the integrated model of organisation for supporting process approach to the statistical production (all as part of current practice) – which requires a lot of time with such a large scale of activities of the Polish statistics.

6. The Integrated Statistical Production Process

The future organization of public statistics should be based on the so-called process approach – using the familiar and successful international solutions, i.e. GSBPM. Among the basic features of the mentioned process approach one can mention:
1. Concentration of entire activities within the organization of statistical public surveys;
2. Adapting a perspective of the “from the outside to the inside” analysis, based on taking client’s expectations as a departure point with regard to a product which is to meet these expectations;
3. Firstly, concentration on relations (so-called “interfaces”) between particular processes and organizations, secondly, looking at the inferior of the organization or the process;
4. Activities (in particular processes) should end with a well-defined (qualitatively and quantitatively) product (half-product or final product);
5. Activity (especially processes) should start after receiving a well-defined (quantitatively and qualitatively) initial product or half-product, so-called “input”.

The presented solutions, are aimed at increasing efficiency of processes of the public statistics organization system while preserving their effectiveness. It seems that the increase of efficiency can be achieved through:

1. Optimization and standardization of labour processes;
2. Specialisation of organisational units;
3. Consolidation and unification of IT solutions;
4. Optimization of the collecting data manner – single time collection of variables added value used simultaneously in many surveys;
5. Maximized use of already used information – e.g. information available in the registers of public administration;
6. Minimization of the use of “costly” methods of data collection – i.e. by respondents;
7. Minimization of data collection social costs – i.e. limiting engagement of respondents.

The concept of the new model of statistical surveys is based on the perspective of work processes. The description of fundamental elements of concept solutions was presented while discussing particular groups of model processes including: specification of needs, designing, construction, data collection, data processing, data analysis, data dissemination and ex-post evaluation.

Organization of surveys of public statistics requires a holistic approach – in particular looking through the prism of a single survey but also through the prism of all surveys and, as was already signalled earlier, drawing attention to the fact that the effect of the entire organization does not have to be simply a sum of particular effects and can be, in a major way, strengthened thanks to the synergy effect.

From author’s analysis it can be inferred that it is possible to significantly simplify the architecture of work processes – i.e. reduction of numerous independent production processes with single processes (Figure 4) – in particular thanks to implementation of task specialization and related competence concentration.

Taking into consideration all observations – including the possibility to secure the proper management of work processes – it seems possible that a certain improvement (and simplification) can be performed on the organizational structure of the organization system of public statistics, where in the place of “silos” one should introduce organizational units with the following task specifications (functional) – i.e. specialization regarding:

- data processing
- data storage
- data analysis
- data dissemination (i.e. customer service)
- organization, coordination and control
- providing IT services.

Activities of the aforementioned “specialized” units (further referred to as key units) require also the support from supporting units – dealing with managing staff, servicing administration and finance. Along with the implementation of specialized units it will be required to appropriately organize them e.g. by creating appropriate profiled teams with precisely established team roles, preparing work and creating action procedures.

Invoking the idea of a holistic approach to the organization of statistical surveys, one should pay attention to the fact that the structure of work processes should be strictly correlated to the so called “production environment” used within the same processes. In the case of the public statistics the production environment is, first and foremost, consisting of the IT solutions used to gather, process, analyse and disseminate data.

The idea of reorganization presented herein is strictly related to the need for change of the production environment of the public statistics surveys organization. It seems that the integration, consolidation and standardization of IT solutions – within one unified production environment – will be an effective method of supporting reorganization,
optimization and standardization of work processes. Additionally, the mentioned consolidation of IT solutions should, in a major way, influence the optimization of development costs of those solutions and their later support.

7. Organizational structure and key tasks

Organizational structure is a relevant element of the proposed Integrated Statistical Production Process which should be coherent with the concept proposed and described above. It is also relevant to enable and support positive effects that it brings. According to the process groups and processes we can distinguish between basic (primary) units:

- Control, coordination and organization Unit
- Data collection Unit
- Data processing Unit
- Data analysis Unit
- Customer service Unit
- IT services Unit

and supporting units (dealing with staff management, administrative and finance management).

Fig. 4. Basic units of the Integrated Statistical Business Process Model

The basis assumption determining the multi-layer structure proposed herein is the concentration on specifically identified uniform tasks for task groups implemented within work processes in the integrated task implementation model followed by focusing – within a particular organization unit – on knowledge, experience and competence necessary to perform these tasks.

One of the main tasks within the Integrated Statistical Business Process Model will be the comprehensive ensuring of efficiency the entire string of activities regarding the exercise of statistical surveys. It is important that the safeguard of integrity of the statistical processes was the unit having appropriate competence allowing it for an efficient coordination of the proper cooperation of all remaining basic organizational units. Therefore the owner of the string of process groups was the organization, control and coordination Unit (i.e. unit responsible both for its entire result as well as for defining and monitoring the measures of its efficiency, entire management of the string of process groups, defining its course etc.).

The fundamental task of the data collection Unit is to gather, on the basis of the necessary preparation, by the data processing Unit different data collections with the use of the most adequate methods of obtaining data. The products of this unit are the data collections gathered for the statistical products group with the appropriate Frames and are transferred to the data processing Units. In relation to the range of implemented tasks this unit was the owner of the process group of data collection.

The fundamental product of the work of the data processing Unit is the micro and macro-data Repository appropriately prepared and fulfilled with data. One of the tasks is to analyse the needs of the data analysis Unit. This need refers to the aggregates and need for resultant micro-data of statistical products groups and processing of the obtained data (i.e. value collections of variables collected for the statistical products Groups) into a collection of data constituting the basis for preparing information and conducting analyses. This unit was the owner of the process group of data analysis.

One of the tasks of the customer service Unit is the preparation of statistical project products on the basis of the information needs identification (i.e. requirement specification regarding the products fulfilling the identified information needs of statistical information recipients). These statistical project products create the needs for statistical products.
The second key task is making statistical information available in the form of final products prepared in the statistical production processes, including their transformation to the selected availability form (printed publication, electronic format). A relevant element of the unit’s activity should be also the promotion of offered public statistics products. Due to the specificity of the implemented tasks by the customer service Unit, it is suggested that it was the owner of the needs specification process group and the data dissemination process group.

The IT services unit was classified as an organizational unit. That was mainly due to the relevant role of IT solution that it delivers and maintains which help in effective support of the activities of remaining units. Ultimately, that activity has an influence on the efficiency of the entire production process of public statistics. The IT services Unit, on the basis of requirements regarding the tools or IT services specified by all remaining units, prepares and implements the IT tools and services. In the context of IT tools construction processes, this unit exercises, adjusts, tests and implements a set of tools and IT services, ensures continuity of work of the entire IT environment supporting the work of the remaining units. This unit is the owner of the IT tool/services designing process.

In order to proceed to the CSO organizational status proposed herein, it is proposed to use the gradual approach, i.e. step-by-step and stretched in time. Such an approach seems justified not only by the scope of changes but also by the fact that the proposed changes will require proper preparation (including motivation) of people engaged in this process. People constitute the fundamental capital of the organization, which, if appropriately organized and directed, determines the success of each organization functioning. People – with their knowledge and willingness to conduct creative operations – shape the organizational culture and formulate action strategies and implement them. The effective use of the proposed competence and the development of new ones constitute a key success factor for the process of proposed organizational changes.

8. Summary

Implementation of the described changes concepts constitutes a strategic decision, bringing benefits in the perspective of many years. Still, as in the case of every organizational transformation it is related to certain costs, therefore making a decision on enacting the Integrated Statistical Production Process should be naturally preceded by a further, more detailed analysis of the benefits and costs. However, comparing the present conditioning, in which the public statistics functions, with the benefits stemming from the use of the proposed solutions, a great potential seems to be lurking in the proposed model. This model seems to eliminate numerous following inefficiencies:

1. Only several connected and integrated IT systems instead of hundreds existing;
2. Only several specialized organizational units, extracted according to task specialization, concentrating specialist competences instead of many “independent” organizations;
3. Instead of using independent and individual approach to the realization of each of the surveys – standardize of realization statistical surveys processes, preparing common methods and the best practices, used for the various stages locally and internationally according to CSPA [4] architecture;
4. Comprehensive look at all conducted surveys – elimination of structural ineffectiveness sources by proper products grouping;
5. Instead of collection data independently for each survey and lack of cooperation for the needs of other surveys – using dedicated system which would support collection data for many surveys;
6. Instead of expensive data, collected from direct surveys, it is suggested increasing obtaining cheaper data – from alternative sources like administrative sources or Big Data and sharing data within the organization and many surveys from different fields;
7. The System of Statistical Metadata in which all metadata of input variables will be gathered will be a tool for inventorying and integrating the variables.

As regards data standardisation and integration, in different countries, an intention to modernise the statistical data is observed, particularly taking into consideration the process of collecting data. The necessity to introduce changes the effect of which will be increase in effectiveness of collecting data when reducing the burden on respondents is stressed also by Eurostat. Particularly, the issue of data
standardisation and integration is emphasised which in the next step will lead to increase in effectiveness of the statistical production. Existence of unjustified differences in the metadata which makes it difficult to integrate collecting data both on national and European levels is universally noticed. Data and metadata should be thus comprehensively reviewed to standardise similar variables which with no great standardisation operations (e.g. by standardising formats, definitions, acquisition time limits, etc.) will be possible to be integrated into one variable used for various types of research areas.

In 2013 Central Statistical Office of Poland has already started implementation at above mentioned concept of the new organization of statistical surveys. Full implementation is expected about 2020.

9. Bibliography


http://www1.unece.org/stat/platform/display/hlgbas


Wprowadzanie ram architektury korporacyjnej w statystyce polskiej

J. DYGASZEWICZ, B. SZAFAŃSKI

Artykuł poświęcony jest modernizacji procesu produkcji statystyki publicznej. Punktem wyjścia do formułowania zasad i metod modernizacji jest ustanowienie ram architektonicznych. Dogłębna analiza wszystkich istotnych aspektów wynikających z holistycznego podejścia do zagadnienia produkcji statystycznej jest podstawą do wyodrębnienia kluczowych zagadnień biznesowych. Potrzeby biznesowe stają się fundamentem do sformułowania spójnych i przejrzystych wytycznych, postulatów, wymagań prawnych i technicznych, zarówno krajowych, jak i międzynarodowych stanowiących podstawę do budowy ram architektonicznych oraz rozwoju rozwiązań informatycznych wykorzystywanych do realizacji zadań statystyki publicznej. Proces konstruowania ram architektonicznych wymaga zastosowania modeli i metod stosowanych w badaniach naukowych, w tym zwłaszcza wynikających z dorobku architektury korporacyjnej oraz modelowania obiektowego narzędziami informatyki.

Słowa kluczowe: ramy architektury korporacyjnej, badania statystyczne, GSBPM, TOGAF.