WOJSKOWA AKADEMIA TECHNICZNA

im. Jarosława Dąbrowskiego

WYDZIAŁ CYBERNETYKI

SUMMARY

OF DOCTORAL DISSERTATION

Models and methods in the process of constructing a framework for IT support to massive statistical surveys

Janusz Dygaszewicz MSc Eng.

Dissertation Supervisor

Prof. Bolesław Szafrański, PhD Eng. WAT

Warsaw 2018

The subject of the study in this doctoral dissertation is the statistical production process used in official statistics, with special regard to the task of IT support of this process. The IT and organizational solutions proposed by me for the IT Census System, whose use in the 2010 census round had fundamentally changed the approach to building and using IT tools supporting statistical surveys, were the basis for this dissertation. The positive results of the censuses became an inspiration to take up a study that, using scientific methods, aimed to generalize, expand and systematize the author's ideas and concepts created for purposes of the above mentioned censuses, with the intention to acquire results, which can be used in the 2020 census round.

After thorough analysis, it has been pointed out herein that the existing flaws and problems of the methodology of conducting statistical surveys have their primary source in the lack of a process approach to statistical production, which has resulted in the creation and maintenance of dispersed, autonomous, inconsistent and ineffective solutions for IT support of statistical surveys. They were apparent, among other things, in the fact that approximately 200 statistical surveys included in the annual Statistical Survey Programme were supported by a comparable number of local IT systems. Considering the above, this dissertation assumed the following:

- use of a state-of-the-art IT solutions in the above mentioned censuses (resulting in an almost total elimination of paper usage),
- use of information resources already at the disposal of public administration,
- an integrated process-based approach to conducting censuses, resulting in a statistical production model different than the one applied so far and a different IT support architecture.

It has been assumed in the dissertation that the achievement of the intended objectives requires introducing a new statistical production model. It is worth noting that the validity of the majority of the ideas compiled by the author of this dissertation in years preceding the 2010 census round, was later confirmed in the Generic Statistical Business Process Model (GSBPM), created by the United Nations Economic Commission for Europe (UNECE).

Considering those, the most important research problems undertaken in this dissertation included:

- an analysis of the current status and the development of a new statistical production model, with special regard to the geospatial component,
- the development of an architectural framework using object-oriented modelling (main topic of the dissertation) and the mathematical modelling (secondary topic). This framework should serve as a basis not only for the development of the IT support of the next census but also as a factor that has a major impact on the process of statistical surveys' reorganization, which has been initiated in Statistics Poland,
- the development of methodical guidelines for using the architectural framework in processes of statistical surveys' IT support design,

- an analysis and discussion on the current IT support architecture and the new architecture compliant with the new architectural framework,
- discussion on the multifaceted comparative analysis (including a cost analysis).

Analytical and construction tasks were carried out in compliance with the corporate architecture principles using object-oriented and mathematical modelling. This approach facilitates the harmonization of IT solutions in Polish official statistics with solutions in other countries, creating a basis for an interoperable international cooperation, especially in the field of reusing data processing models.

In this dissertation, according to my academic and professional achievements so far, I focused mostly on the conceptual-architectural aspects of designing IT support for statistical surveys. However, discussing the usage of the architectural framework I made an attempt to create a general mathematical model, which – provided the necessary data is collected and the work is continued – may have a practical impact on the optimisation of project conclusions, e.g. concerning the choice of census data acquisition or access channels.

As a result of the conducted considerations, in this dissertation I presented the following:

- a new process-based Statistical Production Process Model (MPPS),
- the architectural framework for building and usage of IT support for statistical production conducted according to MPPS,
- methodical guidelines for creation and usage of the architectural framework; (despite the fact that they have been included in the annex due to their volume, the guidelines are an important result of this dissertation),
- discussion on the essence of the transformation of the current end-of-pipe approach, characterized by dedicated separated applications supporting individual research processes, to an integrated process-based approach supporting all research processes through a consistent IT support platform,
- proposal for a technical-systemic environment architecture to be used in the 2020 census round,
- results of a comparative cost analysis.

This dissertation consists of the introduction, five chapters, a summary and three annexes.

Chapter 1 describes the legal and organizational basis for using IT tools and methods in conducting statistical surveys in official statistics. I described the statistical survey term, referring to the mass survey term and census specifics and also presented the necessary definitions, terminology and classifications used in Polish and worldwide official statistics. I pointed out that in numerous Polish and worldwide legal acts the necessity to use IT for statistical surveys is emphasized in order to lessen

the burden on respondents, decrease costs and accelerate the performance of statistical surveys while ensuring high quality results. In this chapter, I also highlighted the role of official statistics in the country's information system. Concluding this chapter, I justified the need to change the current method of conducting statistical surveys, employing an approach built on a process-based statistical production model and complex usage of IT support, designed in accordance with the corporate architecture principles.

In **Chapter 2**, I formulated the research problem and dissertation theses. Basing on the analysis of the manners of creating and applying the IT tools dominating in Polish official statistics, I identified the most common problems and flaws. Considering the above, I determined the research domain, along with goals and the dissertation realization concept.

Chapter 3 contains a thorough analysis of IT solutions dominant in Polish official statistics. The detailed results of this analysis are presented in Annex No. 1 to this dissertation. Results of the analysis prove that it is not possible to create a single, general schema, which consistently describes the structure and process flow of the complicated current statistical production process. In light of this conclusion, creating a new process-based statistical production process model (MPPS) was inevitable. The new model was created with special regard to the geospatial component that has been gaining impact in the contemporary statistical production process.

In **Chapter 4**, I described the method for constructing the architectural framework for IT support of the statistical production based on MPPS. The methodical guidelines for using the developed architectural framework in IT support design and application, due to their extensiveness, were placed in Annex No. 2 to this dissertation. The designed architectural framework is a set of UML diagrams, containing one diagram that presents a complex view of MPPS phases and eight (8) diagrams presenting the structure of conducting each of the phases separated within the MPPS. Descriptions and tables that illustrate the relationships between sub-processes of the statistical production process and data collection objects, which, for the purpose of this dissertation, have been called data repositories, have accompanied the diagrams. It is essential to note at this point that this dissertation also includes conclusions from the mathematical modelling of the statistical survey process have been summarized. This topic concentrated on the possibilities of the stimulation of IT support architecture by the results acquired from the design and study of statistical surveys' mathematical models. The results of the mathematical modelling provided in this dissertation are of a supplementary significance to the main research topic. It is worth noting however, that owing to the inclusion of the mathematical modelling in the management process of IT support design for the MPPS model, its results automatically become the constructional elements for the architectural framework. In the conclusion of this chapter, I described a proposal of integration (in the process of IT support design for statistical production) the results of architectural-constructional and mathematical modelling.

Chapter 5 contains a proposal for IT support architecture for statistical surveys, including censuses, along with a discussion on transforming the current end-of-pipe approach, characterized by dedicated

separated applications supporting individual research processes, to an integrated process-based approach supporting all research processes through a consistent IT support platform. As a result of the designed architectural solutions it was possible to propose a technical-systemic environment for the Population and Housing Census 2021. Proposed solutions were compared with the solutions initially tested in the Population and Housing Census 2011, presenting, among others, the conclusions from the performed cost analysis.

In the **Summary**, I described synthetically, the basic results of the conducted research in the context of the research problem and the dissertation realization concept. I outlined the directions of further research and construction works, depicting the following issues as most promising:

- expanding the research work and the supplementary methodical work aiming at extending the GSBPM with sub-processes using geospatial features, with respect to proposals included in this dissertation. My initiative in this regard was met with high interest, becoming a starting point for international modernization ventures, aiming at a formal extension[i] of the worldwide GSBMP v. 5.0. with geospatial components,
- continuing works on the application of mathematical modelling, with a universal model and method repository along with methodology of their use in processes of planning and conducting statistical surveys (including solving optimisation tasks), as one of their most important results,
- combining the further research-construction work with that ensuing from the development of methods and technologies described presently as Big Data. The essence of this work has to be the creation of models and methods, which will result in reliable statistical data calculated from data, which is unstructured, incomplete and comes from a source whose reliability is hard to be assessed. The possible implementation of results acquired in this domain must be preceded by the legislative changes regulating methods of using alternative sources by official statistics.

An important characteristic of the approach adopted in this dissertation is its conceptual compliance with the recommendations of the international statistical organizations (EUROSTAT, UNECE, UNSD and OECD). Because of the above, the acquired results may have a positive impact not only on changes in the organisation of statistical surveys performed by Polish Statistics, but also on the results of work of the international group of experts, in which I have been representing Polish official statistics, dealing with methods, models and IT tools supporting statistical surveys.

[i] Since 2013, I have been participating in the work of the Modernisation Commission for the Production and Statistical Methods under the supervision of High Level Group with the United Nations Economic Commission for Europe (UNECE) and since 2017 in "Supporting Standards Group on GSBPM", a group of international experts operating within UNECE.