Practice of the Republic of Azerbaijan

on creation and development of statistical metadata system

In globalization period the process of statistical data production, including technologies, methods, and classifications used in production of data always attracts serious interest of users. In certain cases users pay more attention to characteristics of data formation, rather than data. In order to be sure in quality of data they would like to receive "data about statistical data", i.e. metadata.

The requirements and interest of users in this matter are in attention of the majority of statistical organizations that are aware of increasing of these trends in the next future and take into account all new challenges for planning of activities. These challenges do not only mean to be ready to answer these requests, there should be also the comprehensive organization of the processes and system and sustainable effective maintenance of them. In other words, the effective organization of statistical metadata system is necessary, and mainly in interest of statistical offices than users. It may be no accident that creation of metadata system was assigned as task to the State Statistical Committee in the "State Program on development of Official Statistics in the Republic of Azerbaijan in 2013-2017 years", approved by Decree of the President of the Republic of Azerbaijan №2621, dated 21 December 2012.

The "metadata" is a new expression in the country's experience and therefore, in certain cases users perceive metadata as methodology or use this concept in less degree or with caution. Taking into account abovementioned there is necessity in clearly and detailed explanation and promotion of this definition.

In simple phrase, statistical metadata provide information on statistical data. It is referred to final statistical data produced and provided for users. As metadata facilitate explanatory presentation of statistical data, users can obtain detailed information on definition, source, contact and collection method, periodicity, unit of measurement and other. At the same time, statistical offices gain users 'confidence.

Along with detailed explanations, the metadata also cover resources used for collection, processing and dissemination of data, in other words information on specialists and technologies that participate in these processes. So, there is a need to use wider definition of metadata – "statistical metadata system" instead of

“metadata”. Statistical metadata system should be analyzed in frame of 3 components – human resources, processes, and technologies.

During recent years, metadata became component part of the information technologies and there are different structures and experiences in the world that are used for presentation of metadata. Basing on Dublin Core Metadata Initiative these requirements are reflected in “Requirements on creation and management of internet information resources of the state authorities” confirmed by Decision of the Cabinet of Ministers of the Republic of Azerbaijan №189, dated 4 September 2012.

Taking into account abovementioned, the State Statistical Committee contributed to this action by creation a special “Metadata” section in the website. At present, wide range of users can obtain relevance, accurate and complete metadata on indicators and questionnaires at http://www.stat.gov.az/menu/7/.

As indicated above, in wide sense the metadata cover not only data, but also statistical processes and information technologies. Statistical processes include all production processes from specification of needs for a new statistics to archiving of data and metadata and other processes related to statistical information system. It is necessary to document the sequential implementation of these processes, i.e. to create a metadata. The best example of statistical production processes is the Generic Statistical Business Process Model (GSBPM, version 4.0), prepared by the UNECE Secretariat and approved by the METIS Steering Group in April 2009. Referring to this model, the State Statistical Committee carried out standardization of the main stages of statistical activity by means of metadata creation. Depend on characteristics of statistical product the model makes possible to coordinate the activity with human resources and to maintenance the system based on specified technological scheme. For example, in the frame of activity of CPI production, specification of needs, design, building of the production system, as well as processes on data collection, processing, analysis, dissemination, archiving and evaluation stages should be done. This action serves increasing of quality of processes and data. The role of metadata here is to record (in written) these processes and information system that was applied. The creation (writing) of metadata is effective from the standpoint of standardization of statistical production and increasing of efficiency.

Before application of statistical metadata system each statistical organization should describe its essence, goals and functions, as well as stages of implementation, and should plan activities regarding to information system.

The main superiority of creation of statistical metadata system is systematization and documentation of the activities and the job is organized on the base of well-established system, but not on the characteristics of the human resources. For example, in case the employee retires he/she couldn't take the knowledge gained at the job and the new employee organizes the work based on written metadata. This is efficient from standpoint of sustainability of organization and reduces the risk of knowledge losses. Moreover, in the result of staff's inexperience or poor memory during organization of the survey, the process that would be overlooked and errors that would arise are minimized and as a result, quality of data is increased.

According to practice, metadata is of great importance on management of projects, solving of organizational matters, and strategic planning, efficient human resources management and etc.

The principal users of statistical metadata are respondents, representatives of government, mass media, and other users of statistical data at national and international levels, as well as the users within the statistical office - leadership, statisticians, methodologists, creators and members of information system.

By means of metadata the head of the organization controls and manages statistical processes in the statistical information system and activities of the specialists and makes right decisions. For example, the head of the organization supervises the work of specialists participated in preparation of statistical data, for what stage and for what job they are responsible in statistical production, what process was not implemented and etc. and as a result, ample opportunities for right strategic management are being gained.

Metadata is also useful for statisticians, methodologists, creators and users of information system, so by applying to metadata of previous surveys they are familiarized with experience, obtain information on register, classifications, standards, methods, software used and increase the knowledge, and organize own activity properly.

Respondents, representatives of government, mass media and other users of statistical data at national and international levels use metadata on questionnaires and final outputs.

Before giving information on priorities related with standards in the field of common metadata framework that is being applied by the SSC of Azerbaijan or will be applied in future, it should be made a point that decision on creation of statistical

metadata system in the Committee was made in 2011 taking into account importance of abovementioned and challenges as well as recommendations of Global Assessment of the National Statistical System of Azerbaijan, that was carried out in 2010 by UNECE, EU Statistical Bureau and EFTA, on creation of quality management and metadata systems. For this purpose, experiences of other countries on metadata framework have been learned and the current situation has been analyzed in detail. One of the achievements of the Committee in this direction was establishment of the separate structural division, the main activities of which related with metadata. The "Action Plan for 2012-2013 years on creation of quality management and metadata systems in the State Statistical Committee" has been confirmed by the Decision of the Board. A common structure and software is prepared for official statistical questionnaires and indicators; the metadata more than 1000 statistical indicators and questionnaires are placed in website and all available metadata are presented in different formats. Based on internationally accepted GSBPM, 9 stages and 47 sub-processes of statistical production on main areas of activities are prepared and the maintenance of records is started.

Proper and long-term organization of harmonization between data and metadata of the international organizations and local statistics improves using of the international standards. The SSC attaches significance to application of the international standards on creation and development of the metadata system by taking into account national peculiarities. According to analysis of specialists of the Committee, there is no a country of best practice regarding to ideally organization of statistical metadata system. The specialists of most countries, including statisticians of Azerbaijan carried out own work based on existing standards of statistical metadata, statistical concepts, technical standards, models, and statistical experience, methodological guidelines and recommendations.

The metadata system in official statistics begins from application of statistical classifications and registers. The SSC of Azerbaijan gained serious success in this field, so national versions of all internationally applied classifications have been prepared by the Committee. 24 classifications fully conforming to international and 6 national statistical classifications based on country's legislation have been prepared and used in statistical offices as well as in country's territory as common standards. At the same time, classifications of different fields have been carried out with the purpose of using by statistical offices. It gives possibility to manage metadata based on classifications and integrate them to information system. The same refers to register. Taking into account the fact that the registers play a

significant role in data management, and integration of input and output data, the managing of database, organization of observations, and summarization of data and preparation of users' tables is carried out based on these registers. Registers in Azerbaijan that were created for statistical purposes since 1995 have been completely formed on the base of statistical units' single register.

Harmonization of one of the key objects of the statistical metadata, such as statistical concepts and terms is considering the unification of statistical products' name and metadata. Fully realization of works in this field is very hard process. So, there is necessity of a long time and human resources for direct integration of provided data with the output data, coordination of terms used in different processes and linking them with metadata. For example, the statistical product "industrial production index" used by users is based on many columns and lines of different questionnaires. The number of these columns and lines is calculated by ten thousands, but statistical product in this example equals to one. At the same time, the question "the volume of product you produced" for respondent means "volume of produced product" for user and application of e-connection between them and metadata for all processes is one of the main tasks. The tasks also include glossary that should be prepared for indicators and metadata.

Formation of metadata system requires application of technical standards. As stated above, for both statistical and non-statistical sphere, the metadata are disseminated in e-resources according to Dublin Core metadata standard. Dublin Core metadata standard is considered as the most popular universal metadata system from internet resources using standpoint. This standard simplifies access of producers and users of statistical data to global information network by means of small amount. Dublin Core metadata standard is also used for preparation of metadata catalogue and indicating the existing data sources. Moreover, in international practice Statistical Data and Metadata eXchange standards (SDMX) are more preferable for statistical data and metadata exchange. Recommended by Eurostat and consisted of 21 concepts and they sub concepts, these standards are prepared for statistical purposes and will be the main standards in the next future. The eXtensible Markup Language (XML) format is used in these standards for description and exchange of metadata. Based on SDMX standards, the countries could easily integrate all disseminated data with its own metadata system; also the international organizations obtain single access to the countries' metadata. Along with Dublin Core and SDMX, one of the best examples in the field of documentation of statistical processes is Data Documentation Initiatives (DDI). DDI standard is mainly used for internal use,

Dublin Core and SDMX- processes is foreseen for user. This standard is of XML format and makes possible integration of processes by web technologies.

Users of statistical data emphasise quality aspects of obtained data and metadata and require its relevance, accuracy, clarity, coherence, comparability, timeliness. The State Statistical Committee declared its purpose and policy with regard to quality assurance: confirm abovementioned action plan, establish Council for better coordination of quality issues and assign quality representative. As a result, recording of documented procedures and preparation of schemes of processes by activities was reached, and application of quality management and metadata systems was started. According to gained experience, quality assurance should be developed in parallel with metadata system. It is necessary to note that according to UN recommendations, the "Managing metadata" is one of 19 core principles of quality assurance and the Committee pays special attention to this matter and uses the best practices and standards, accepted at international level for effective metadata management.

As stated above, the quality management of statistical production in SSC is carried out based on national version of GSBPM and now it is realised improvements that consider the coverage of more detailed activities. For preparation of quality reports the national "Standards and recommendations for reports on quality assessment" were confirmed based on the "Law on Official Statistics" of the Republic of Azerbaijan and other normative documents, as well as the "Code of Practice of the State Statistical Authorities" and "ESS Standard for Quality Reports". Preparation of quality reports, the main object of which are statistical processes and its outputs, mainly depends on quality of metadata and the following aspects should be reflected here: relevance, accuracy, timeliness, accessibility, coherence and comparability, cost-effectiveness, assessment of respondent burden and other aspects. Quality reports will enable to analyze strengths and weakness of statistical production and facilitate comparison between statistical producers. 23 quality reports on main statistical products have been discussed and confirmed at the meeting of the Scientific-methodological Council on 27 May 2013.

- Calculation of GDP indicator;
- One-time statistical survey on studying of structure of expenditures for production of goods, provision with services;
- Statistical observation of municipal authority budget's execution;

- Statistical observation of activities of insurers and legal entity - insurance brokers' ;
- Calculation of industry production index;
- Calculation of capital investment indicator;
- Compilation of energy balance;
- Livestock statistics;
- Crop production statistics;
- Statistics on air pollutant emissions;
- Statistics on cargo transportation in transport sector;
- Tourism statistics;
- Crime statistics;
- Internal trade statistics;
- Foreign trade statistics;
- PPI;
- Price index of imported and exported goods;
- Cargo transportation tariff index;
- CPI;
- Estimation of the population size;
- Households budget's statistics;
- Number of employees and calculation of wage;
- Food balances.

Guidelines and recommendations of UNECE, IMF, OECD and other international organizations related with the requirements on metadata dissemination have been studied by the Committee and corresponding recommendations have been prepared and applied taking into account national peculiarities.

Studying of all available standards, models, recommendations and best practises enabled the SSC to provide purposeful application of statistical metadata system and as a result let the users, including the Committee, benefit from effective managed metadata.

All abovementioned permits to note that strategic management of statistical metadata system in the SSC has been integrated into general strategic management of the Committee. At present, realization of each stage of management of statistical metadata system is being continued successfully based on support of the Committee leadership, and by involvement all working staff - statisticians, methodologists, IT specialists, and etc.