

Topic (iii): Metadata and quality

METADATA AND QUALITY

Submitted by the National Statistical Committee of the Kyrgyz Republic ¹

Contributed paper

I. INTRODUCTION

1. The National Statistical Committee (NSC) of the Kyrgyz Republic measures data quality and submission methods to statistical information users. The importance of these activities is growing with the expansion of the use of Internet. In this report the necessity of the development of a standard for statistical metadata and data is considered. The situation is explored, in links with the International Monetary Fund's General Data Dissemination System (IMF GDDS). Arising problems are identified and directions for future work are indicated.

II. DISSEMINATED DATA QUALITY

II.1 Metadata

2. Statistical data quality defines its availability for use. An accepted definition of data quality does not exist and, therefore, it is difficult to make an evaluation. Statistical data preparation and dissemination methods used at present include a mechanism for statistical data observation and correct interpretation of statistical information received through requests performed using metadata.

3. The first step in metadata development in the Kyrgyz Republic was their preparation in accordance with IMF GDDS. Using GDDS as the statistics development base sets up conditions for timely receiving and dissemination of reliable, exhaustive and available economic, financial and social-demographic data. For the purposes of participation in GDDS, the following metadata have been prepared through the joint efforts of IMF and four establishments of the Kyrgyz Republic:

- ?? National Statistical Committee: real (national accounts, national accounts aggregates, production indices, price indices, labour market indicators) and external (export/import of commodities) sectors; social-demographic data (population, health, education, poverty);
- ?? Ministry of finance: budget unit (central government operations, central government debt, central government budget aggregates, external debt and external debt service schedule);
- ?? National Bank: financial (review of wide money supply, wide money supply and credit aggregates, National Bank aggregates, rate of interest) and external (balance of payment, balance of payment aggregates, international reserves, exchange rates) sectors;
State Commission on Securities Market: share market included into financial sector.

¹ Prepared by Zarylbek Kudabaev.

4. Metadata on data reliability and access to the public with exposition of acting policy and practice have been prepared by each of the above-mentioned offices.

5. How can measurement of data quality be provided to the statistical information user? The approach of the National Statistical Committee to this question in accordance with GDDS is carried out as described below.

6. Firstly, it is a submission of detailed information on definitions, classifications, data sources, methods of compilation and use of international standards included in the metadata in a comprehensive framework. The information provided in the metadata does not include a specific quality assessment but does allow users good results with respect to the data quality aspects.

7. Secondly, with reference to data categories and indicators, the quality parameter envisages two methods for specific assessment: (i) to provide documentation on methodology and sources used during the data preparation, showing the order of data compilation and data sources; and (ii) to disseminate data on component specification, checking with adjacent data and statistical bases, thus providing statistical cross checks and ensuring validity.

8. In our opinion these methods are more acceptable. Users, who are provided with statistical assessments only, are usually not able to evaluate their accuracy on the basis of their numerical value. Availability of documentation on methodology and sources which are based on statistics is the key factor which allows users to assess the data advantages and amendments and to judge what is essential in international methodological recommendations.

9. So, in many cases the issue of statistical data in the Kyrgyz Republic is accompanied by technical and brief comments of a methodological character on statistical indicator components and contains information on sources used during data preparation. More detailed explanations on methodology contain annual and thematic publications.

10. Data are indicated during the publication of preliminary data. Information is also provided in the case of data revision. Timely information on vital changes in methodology is not usually given but, where there are such changes, data issue is accompanied by short explanations.

11. Specification of dissemination data components permits cross-checking and has self-dependent value for detailed analysis. An adequate volume of detailed data published on the level below the aggregates, particularly in combination with documentation, gives the user an opportunity to verify data. The degree of comparability is enhanced due to data integration on a general basis, as a System of National Accounts, and checked via the comparability of coherent lines. It permits observation of the same phenomenon under the divergent angles. For example, the user has access to the information by main categories of final use, share of main economic branches to total GDP, regional volumes of gross product by branches and components of final use by regions of the Republic, external operations accounts in the annual publication on National Accounts of the Kyrgyz Republic. Indices by main types of activity and more important individual products are published by both industrial products physical volume index and an aggregated index. The data on physical volume index are grouped both separately by types of products and by type of agricultural activity, and as a whole by agriculture, which permits the user to compare and control the data. The consumer price index is disseminated separated into goods and groups, permitting verification. Detailed geographical information (country origin for import; country destination for export) is provided for the purpose of comparison on export and import of goods. Information on goods is published by codes of Goods Nomenclature for Foreign Trade (TNVED) and Standard International Trade Classification (SITC).

12. Comparisons and checking within the Republic cover all data received from different sources, for example: employment data are available in labour resources statistics, population census and households survey.

13 In perspective, the National Statistical Committee is planning to issue the Methodological Conditions on Statistics, which will contain activities of a methodological character developed and used by the statistical office of the Republic, and will provide additional information on the methodological basis of statistical data preparation and dissemination.

14. The main quality element is provision of timeliness. This is restricted to the availability of deadlines of data dissemination and observance of these deadlines. The planning of statistical activity is implemented for the purposes of provision of qualitative statistical information to different groups of users. The program of statistical activities is approved annually by the Government of the Kyrgyz Republic and, in accordance with this program, the National Statistical Committee implements the main activity. International and interdepartmental information commitments and fundamental principles of official statistics are considered during its compilation. The program defines the list of information, its periodicity and deadlines of information submission to external users, the circle of executors.

15. Since March 1998 the National Statistical Committee has implemented a new publication policy which envisages publication of statistical information and its dissemination according to the publications plan. The publications plan is carried out annually; it encompasses thematic publications and the web-site of NSC.

16. Users of statistical data must be provided with information for the improvement and assessment of quality. The Program of statistics development in the Kyrgyz Republic for 2001-2005 was officially approved by decree of the Government of the Kyrgyz Republic in December 2000. It includes short- and long-term plans for updating statistics and improving data quality. The main activities of the program have been included into the metadata of the Kyrgyz Republic as plans for improvements.

17. For the purposes of qualitative and timeliness implementation of the above-mentioned plans for improvement, activities concerning their realization have been prepared by Ministries and Departments participating in GDDS and approved by GDDS National Coordinator of the Kyrgyz Republic. NCS implements coordination functions on GDDS and summarizes the results of implementation of these activities quarterly.

18. Metadata, as a main channel of submission of information on data quality, require improvement. In this connection, Kyrgyz statisticians at present face the challenge of creating a central system of metadata which covers not only indicators regulated by IMF GDDS but which completely corresponds to the statistical database providing their continuous updating.

II.2 Internet and its use for integrating data quality with the data

19. Channels of provision users with statistical information include Internet. NSC has an opportunity to use Internet for:

- ?? reception/transmission of E-mail;
- ?? statistical publications dissemination;
- ?? statistical data dissemination;
- ?? data collection;
- ?? communication with users, etc.

20. NSC is on the list of ministries, departments and establishments connected with the Information System of State Administration. This system provides the delivery of E-mail to many ministries, departments, establishments and their subdivisions on the territory of the Kyrgyz Republic from allocated communication channels. At present, this system is in the stage of development. There is only one communication channel per regional state statistical office. Delivery of statistical information through allocated communication channels provides reliability, quality and timeliness of correspondence delivery. Other regional state statistical offices are equipped with modems for transmission of statistical

information via e-mail in the Main Computing Centre of NSC on communication channels (by switchboard).

21. Since March 1998, the web-site of NSC (www.nsk.bishkek.su) is on Internet and contains general information and data on real, budget, finance, external and social sectors, plus files in Excel from the NSC database. Information on the main social-economic indicators of the Republic is updated monthly. Users can receive metadata in Russian and English via the Internet. A new section "Methodological issues" was launched; it contains activities of a methodological character developed by NSC and the Statistical Research Institute.

22. Since introducing the Web-site we have received many requests from libraries, students, citizens and different organizations, including international organizations. We are planning to introduce request statistics for studying the demands of users.

23. With the participation of the IMF resident adviser, a system of statistical indicators was developed and introduced as a centralized database for preparing publications. It contains more than 300 statistical tables. Further, we plan to extend the database to include all processed statistical data of NSC.

24. But there are situations where users need information which is not in the database. We have a primary database but does not really conform. In the future, it is planned to provide user access to common information – statistical database via Internet - and to launch a special program which creates tables on request by the user, and to satisfy the increasing demand.

25. The information on the Web-site of NSC is of a methodological character, describing the methods used to prepare and disseminate data, thus permitting users to interpret not only the data but also their quality. The level of timeliness of information provision is increased. Free acquisition of the metadata is provided and a more acceptable integration of the information is implemented.

26. Internet is of great importance for cooperation with international organizations and statistical offices of other countries, both in using statistical data and metadata. To enhance efficient use of the metadata, references to more detailed metadata in national Web-sites can be included into the metadata in Web-sites of international organizations. For example, the metadata on the IMF Notice Board on Dissemination Standards.

27. Generally, the influence of Internet on the integration of data quality with data has not been looked at in-depth by statisticians of the Kyrgyz Republic. However, we can foresee the shortcomings of a survey via Internet. Not all the observed units have access to Internet. So, during the conducting of households surveys, this shortcoming can influence all basic survey components, for example: differentiation, sex, educational level and it also influences data quality. In the case of enterprise surveys, this problem is less acute. The use of net services for information collection is atill at the development stage.

28. One of the problems of data collection via Internet is the use of unified terminology in the field of metadata and the creation of catalogues as the basis for statistical metadata bases. Taking the marketing and new software available into account, we must develop new approaches for electronic data collection directly from observed units.

III. STANDARDS FOR METADATA OF DATA QUALITY

29. Statisticians of the Kyrgyz Republic draw attention to the correspondence to international standards. It gives an opportunity to conduct international comparisons and analysis of the economic situation in the country for all interested parties and it permits their independence from the government bodies, protecting them from manipulation with conceptual situations.

30. The term “International standards” includes concepts, classifications, system of national accounts and other statistical products in the different fields of activity. In this connection, it is necessary to refer to the standards for metadata on main statistical lines of IMF: General Data Dissemination System (GDDS) for all countries and Special Data Dissemination Standard (SDDS) for the most developed countries. Metadata within the framework of these standards characterize the quality of disseminating data.

31. At present, the Kyrgyz Republic uses GDDS as the standard of statistical data dissemination. The next step is the adoption of IMF SDDS and statistical requirements of the European Union.

32. Metadata include all industrial processes beginning with data collection and finishing with its dissemination. Users of statistical information of all categories need the metadata for implementing the following functions:

- ?? to set up the availability of explicit data;
- ?? to obtain support during the information search;
- ?? to receive the information on data sources, definitions, classifications, methods and methodology of compilation and use of international standards;
- ?? information interpretation;
- ?? information dissemination;
- ?? following information processing.

33. Metadata are characterized by considerable differences depending on the country and applied methodology. Relations of statistical offices with users show the importance of development of a common basis for data quality definition, for the purposes of more effective assessment of comparability of disseminating statistical information at national and international levels.

IV. CONCLUSION

34. We believe that a comprehensive system of data quality assessment must be used in the interest of national statistical offices. All requests regarding information provision to external users, on either an obligatory or a voluntary basis, must answered with information from the same statistical database. Information on data quality should be analyzed and include not only quantitative information but also information on the processes and procedures used which allow interested parties to define a level of data quality.

35. Web technologies give an opportunity for increasing the quality of official statistics. The use of Internet at national and international levels demands efforts in standards development. Standardized concepts of statistical data and metadata should be one of the main draft conditions to meet expectations of users concerning statistical data quality.