STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working paper No. 31
English and Russian

CONFERENCE OF EUROPEAN STATISTICIANS

UNECE Work Session on Statistical Data Editing
(27 – 29 May 2002, Helsinki, Finland)

Topic (iv): Impact of new technologies on statistical data editing

INFLUENCE OF NEW INFORMATION TECHNOLOGIES ON STATISTICAL DATA EDITING

Contributed paper

Submitted by the National Statistical Committee, Kyrgyzstan[1]

I. INSTITUTIONAL BACKGROUND

1. Kyrgyz statistics today are the product of quite an effective information system working on an independent basis for the collection and processing of statistical data.

2. The principal objective of the National Statistical Committee of the Kyrgyz Republic is to offer information services of high quality to users of statistical data.

3. The National Statistical Committee and its regional organs, the Main Computing Centre and subordinate bodies constitute a unified system for the supply of official statistics.

4. The Main Computing Centre implements the National Statistical Committee's common technical and technological policy for the collection, transmission, automated processing, collation and storage of statistical data, employing modern technologies and making wide use of electronic, printing and office automation equipment and communications.

5. The promotion of computerization, which involves creating a modern information infrastructure and building and using information systems, resources, technologies and means of supporting them, is a key to socio-economic progress in the country and therefore one of the highest priorities for the development of the statistical system.

6. A major goal of the "Programme for the development of statistics in the Kyrgyz Republic over the period 2001-2005", which was adopted at governmental level, is to ensure the comprehensive development of the statistical information system based on using modern information technologies for the collection, processing and transmission of statistical data at all levels, and this means fully equipping and upgrading the system's technical base with modern hardware and software.

II. USE OF INFORMATION TECHNOLOGIES

7. The National Statistical Committee's central office and the Main Computing Centre have 174 workstations linked to a local area network. Computers with Pentium or Pentium II processors are used at the workstations.

8. The local area network is basically structured on the commutating segments principle. It is designed to provide simultaneous access to an unlimited number of users.

---

[1] Prepared by Z.I. Kudabaev, Chairman.

9. The software products employed in the National Statistical Committee system include MS DOS, Windows 95/98/2000/NT; MS Office 97/2000 (Word, Excel, Access); Clipper and Delphi database management systems; Visual Basic and SQL programming languages; and modern text and graphics editors.

10. Sets of electronic data processing (EDP) software were designed using Clipper 5.x, FoxPro and Delphi programs. Most of them were created with the Clipper software.

11. For data capture and processing, the National Statistical Committee now has a database server running under the Windows NT operating system, with the Microsoft SQL Server relational database management system; a file server based on the Novell network operating system; and a mail server with Microsoft Exchange software.

12. The local area network is connected to the Internet, but only a limited number of users have access to it.

13. Information can be exchanged by electronic mail with all regional statistical administrations. The central office and regional services providing official statistics are connected to the Kyrgyz Republic's Internet computer network. A National Statistical Committee Web page has been developed as part of the Internet system and the statistical information which it provides is constantly updated. The National Statistical Committee system is gradually being incorporated into the State Management Information System.

III. DATA EDITING SOFTWARE

14. The direction of information flows in the National Statistical Committee system is determined by the hierarchical organizational structure and technology adopted for data entry and processing (editing).

15. Because of rapidly changing statistical data requirements, the shortage of funding available to acquire standard data editing software, and the inadequate "knowledge base" in this field, the work of entering and editing statistical data is done with proprietary software.

16. The Main Computing Centre's programmers have produced more than 150 sets of software, and these are being used successfully at both national and regional level.

17. The software is now being developed mainly using the Clipper database management system, enabling independent modules to be created for downloading with limited computer capacity, which is important in view of the PCs currently available at regional level, and also allowing Clipper dbf files to be used for further processing with MS Office software packages.

18. The software packages designed by the Main Computing Centre's specialists make it possible not only to edit the primary data, but also to implement a number of techniques to verify carefully the agreement and comparability of the data, to analyze the aggregated data obtained in time series, to scrutinize the data for consistency and to verify logically related statistical indicators. The primary data entered can also be checked and edited with reference to administrative data, e.g. reporting by enterprises on industrial injuries and records of inquiries into industrial accidents supplied by the Federation of Trade Unions, sex and cause of death, age and cause of death, a woman's age and child's birth order, etc. The end result of this phase is "net" data ready for the production of summary final tables.

19. The software packages also provide methods (formulae) for computing a number of relative indicators meeting a given criterion. The tables are automatically recalculated when the data is edited or when the initial data is changed.

20. The monthly database of customs declarations of goods is automatically edited for any discrepancies in the data having regard to:

- the value of a good indicated in a customs declaration in the currency defined by contract and the rate of exchange;
- the average export/import price of the goods and the monthly price handbook;
- the enterprises' names and codes and data of the Unified State Register of Statistical Units;
- the product codes of the CIS Commodity Nomenclature of Foreign Economic Activities and units of measurement, including weight and other measures;
- the justification for exporting and importing certain kinds of goods.

21. The local area network now established has solved the problem of exchanging information needed between sections of the National Statistical Committee and the Main Computing Centre, and has contributed to improving the overall quality and speed of handling of statistical data.

22. Standard Excel, Access and SPSS programs are used for editing highly generalized data.

23. SPSS is used in editing intermediate control data and to detect errors during the production of summary tables. It is used as a table-oriented editor for defining, entering, editing and presenting data, and also for setting out the results obtained in graphical form.

24. Excel electronic tables are used for certain calculations necessary both for economic analysis and for compiling time series.

25. Excel is also used to perform SNA calculations. Primary data obtained from the annual database prepared on the basis of returns from enterprises and organizations is incorporated (in dbf format) into sectoral "vectors" and sent for greater analysis and editing to the sectoral departments, which in turn use Excel to perform further calculations, while keeping the structure of the tables produced and formulae used in the calculations. During analysis of the aggregated data, moreover, it is possible to edit incorrect enterprise data, both in the primary reporting of enterprises, and at any stage of the calculations.

26. The "vector method" is also employed in Excel to obtain information about foreign trade. Using "vectors" helps more quickly, and at minimal cost, to provide the information needed on the republic's foreign trade, with respect to the country/commodity and commodity/country distribution and its structure and quarterly and annual movements, and also detail required for the sections and groupings of international classifications (CIS Commodity Nomenclature of Foreign Economic Activities, SITC, CPA).

27. The EUROTRACE Programme is used to obtain monthly updated information on foreign economic activities. This helps to prepare the various statistical tables quickly for analysis and for geographical and other comparisons for the main product categories, and to compile time series (monthly, quarterly, yearly).

28. Delphi and SQL-server software products were used to process material from the 1999 population census. New information-processing technology with client-server architecture was introduced for data entry and adjustment and tabulations, thus making information from the population census database available simultaneously to a number of users.

IV. DEVELOPMENT ISSUES

29. The introduction of standard software and hardware helps firstly to develop the statistical infrastructure and to address major data collection and processing problems.

30. One of the most important goals in this respect is effective implementation of a common policy concerning the methodology, equipment and techniques required to utilize new computer technologies and programs.

31. Under the “Programme for the development of statistics in the Kyrgyz Republic”, arrangements have been made to set up an integrated database - a system making it possible to apply the common policy in order to improve the quality of statistical information and respond quickly to users’ requests.

32. The utilization of modern software for the integrated database will help in creating a user interface that meets present-day requirements and ensures a harmonious changeover to other MS Windows applications (Word, Excel), thus affording greater freedom in the editing and preparation of the tables to be produced, allowing for the data obtained to be presented graphically, etc.

33. Use of common data editing and data entry systems will improve efficiency in the supply of high-quality, timely and accessible statistical information.

34. Countries with transition economies need to know more about the conceptual processes involved in data editing and to study in greater depth the procedures for editing and the methodological, regulatory and instructional material available in this field.