

ICT DEVELOPMENT STRATEGY
2006 – 2010

Belgrade, 23 September 2005

Preface

The collection, processing, analysis and dissemination of statistical data are the principal tasks of official statistics. While carrying out these tasks, it is required to observe the needs of the statistical data users, i.e. the objectives of statistics that are defined by the program of statistical surveys. The starting points for this document are the Act on statistics, the Master plan of statistical development and harmonization and the Strategy papers of the Statistical Office of the Republic of Serbia, 2006-2008, which define further development of information support to the processes and procedures of statistical surveys.

Regarding the substantial changes in methodology, execution and coverage of the statistical surveys that are envisaged for the forthcoming period, as well as in reference to the generally increased demands for statistical data and additional requirements relative to the quality, timeliness and availability of data, the strategy of *ICT* (information communication technology) application is expected to provide:

- efficient coverage, good-quality control and reasonable and reliable data storage
- maximum availability of data within the authorization of users
- high productivity of data processing
- modern solutions in data dissemination
- simple and efficient operations for all users.

Therefore, important objectives of the *ICT* department are:

- development of the applications for the processing and usage of statistical data, support in applying special software for statistical purposes
- creation and maintenance of data bases and access to bases
- hardware maintenance and implementation and maintenance of basic software
- network maintenance
- data protection and archiving
- support to dissemination and statistical publications development, special support to electronic form, with website being the priority
- implementation of internet (intranet) technology
- to monitor *ICT* development and to provide the users (statisticians and IT professionals) with education and training to apply the same
- to create a powerful platform for data and applications processing and to render various IT services, with respective valuation of input resources.

Data input and editing

The data input and editing for the statistical surveys that are found to be suitable for distributed mode of data preparation will be carried out in regional offices, or directly in report units by using electronic questionnaires. For the other statistical surveys, the Data entry section at SORS will be in charge.

To create the applications for data input and editing, **Visual Basic 6** programming language will be used, with gradual introduction of **PHP** (*Hypertext Preprocessor*) and **VB.NET** (*Visual Basic.net*) for making web applications.

We shall consider the cost-effectiveness of installing optical readers for data input and editing, which are expected to reduce manual operations needed for data input and to facilitate the processing of surveys. However our specific requirements are to be regarded (mostly distributed input, variety of questionnaires, coding of certain textual notes, etc.)

The SQL Server database will be used for data input and editing. The distributed *SQL Server* databases at the regional offices will be connected by replication with the main *SQL Server* database at the SORS, in order to automatically transfer data to the Office.

After the data input and editing have been completed at the regional offices, the data will be stored with the SORS, in the *SQL Server* database, whereupon, depending on the statistical survey executed, the data collected will be possibly:

- kept at the PC platform for further processing
- additionally be corrected at the PC platform and then forwarded to *host* (IBM z800, with z/OS operating system), into a DB2 database, where their further processing will go on
- transferred to host, into a DB2 database, where their further processing will go on.

In order to facilitate maintenance and the introduction of standards for the creation of *SQL Server* databases and for applications for data input and editing, it is necessary **to provide a general tool** for creating databases on the *SQL Server* and applications for data input and editing.

Applying standard procedures within this application is required in order to provide the archiving and protection of the *SQL Server* databases and the programs for data input and editing.

The back up copy for all *SQL Server* databases and for the programs for data input and editing will be stored on one server, since this is the way to provide security and protection in cases of possible PC defects.

In order to provide constant connections of high speed and broadband, in accordance with new technologies, the links (connections) with the Belgrade Office will be established for all regional offices and due to specific data processing in Vojvodina, direct connections between their regional offices and the Novi Sad Office will be provided.

These connections will make a network with all regional offices included and a unique closed statistical network, featuring high-grade protection, will be offered.

Databases

In the forthcoming period, relational databases will be used more extensively in regular statistical production. Gradually we shall abandon *RAPID* (relational database developed with the Canadian statistics in the 20th century eighties), *GODAR* and *GODARTAB* (program systems developed with SORS in late 20th century eighties, envisaged for the management of statistical data processing) and transfer to *DB2* (host) and *SQL Server* (PC platform).

The criteria for the selection of certain platform and *DBMS* (database management system) will include: survey coverage, data volume, dynamics, data availability time and series availability.

Regarding the contents, statistical databases are classified as **microdatabases** (individual data), **macrodatabases** (aggregate and calculated data) and **metadata bases** (data about data). According to the place within the statistics production process, databases are differed as **input** and **output** bases.

Input bases are the bases of individual data suitable for checking and correction of records. It should be noted that the demands for data organized in this way are time-limited to the period when the processing of a certain time point of a statistical survey is in progress. Upon the processing having been completed, the data will be archived and the final results will be stored in macrodatabases for permanent usage (table-presentations, ad hoc inquiries, dissemination).

Output bases are the bases of aggregated data, adjusted to the requirements of **providing prompt information for statistical data users, preparing publications and data dissemination**. For these bases it is important to select the right contents and the adequate aggregation level, in accordance with the survey and the data users demands.

The period of individual data storage in tables in a relational database will be defined along with the introduction of a statistical survey and the data character and the very survey will affect it. When this period has elapsed, **the data will be archived applying standard procedures**. On the other hand, standard procedures will be developed and they will provide that the archived individual data can be quickly and easily restored to tables for re-processing or additional processing on request.

Macrodatabase ought to be well designed and convenient for obtaining all required information, as well as for cross-comparison of data of various statistical surveys and providing time (and other) series. A well organized and well maintained **base of statistical registers, codes and nomenclatures** will make a prerequisite segment of this base. The issue of **metadata base** ought to be developed as well, starting from the basic and necessary contents that will be gradually supplemented, according to (current) requirements and circumstances.

For microdatabases, both platforms are suitable: *Mainframe* and *PC*. For every specific survey the choice will depend on the criteria mentioned. The macrodatabase will be on the *Mainframe* and if necessary, due to the convenience of certain tools that are applied for dissemination or statistical analysis and that are connected to PC

platform, a SQL Server base of smaller volume (dissemination base) will be created as a part of the main macrodatabase. The **Host computer offers possibility of concurrent operations for a large number of users and simple control of access to data, as well as protection and security in cases of casual data damage or misuse.**

In future development of statistical survey processing, relational databases will be used as the main way of computerized data organization. Once the processing has been completed, the data will be moved from the databases and archived, provided their further processing is no longer needed or if it is estimated that the processing would not be required for a long period of time. In case the data are restored from the archive, they ought to be placed back again in databases, for as long as they are needed. When macrodata are concerned, each survey provides a certain sum of results that are to be stored in macrodatabase. The only exceptions are some ad hoc surveys, which are carried out as single surveys due to some rather specific output data that only a small number of customers are interested for.

The platform selection (*PC* or *host*) for individual-data bases will be done once all aspects of a survey are considered; however it is difficult to distribute surveys in advance. For a survey where the data are entered in regional offices and processed within a certain period of time, without necessarily being related to the outputs of other surveys, the processing will be carried out on the PC platform only. Otherwise, i.e. if the correct data are provided from some other sources (excluding statistics) and no corrections are to be made, the complete processing will be carried out on the *host*. In all other cases, the both platforms will be used.

To have access to databases, we need well-defined user software, which would provide for possible access from the both platforms. The software will be developed mainly by using PC tools, while the programs ought to be written transparent so that it is of no relevance in which database the data are stored. However PL/I should not be completely excluded, especially in case when already existing programs, or parts thereof can be used.

Data security and protection

The key requirements that the Office adopted for its information processes development are the **security of data and processes** as well as their high-grade **availability** to users.

The existing software solutions (and the software we purchased) offer very good data protection on the *host* platform. Here included is **RACF** (Resource Access Control Facility), the system software that has been tested in production and that has been among the best packages of the kind for many years. It is important that this software is well integrated in its developing and production environment, and carries out selective **control of access to data and processes** and is approved resistant to external intrusions, through the network.

Along with *RACF*, the **DB2 subsystem**, where the main amount of data is to be stored, also possesses a sophisticated system of authorization checking of access permissions to the data in tables and other objects. The point is that these two packages are very well mutually adjusted in their 'charges' to check and to cover the complete part of the production on the *host*.

It is also necessary to ensure a similar degree of **protection** on the **PC platform** and the Office **network**. In order to achieve this, we will introduce **precise standards**, rules and the policy of operation on the PC platform, as well as **a clear structure of privileges and charges** for administrators, programmers and *SQL Server* database users.

PC platform features a system of data protection that is presently used on servers and certain workstations. **The protection system** and the database of users will be centralized and the SORS network will be transferred to the domain structure of an active directory that provides for control of all network resources (user orders, network connections, computers, printers, etc.) from a single computer.

SQL Server database features its own system of **authorization and checking access privileges** to data in tables and other objects. *SQL Server* databases can also be protected by a combined system of *WINDOWS-SQL* protection.

The **network** workstations and servers are presently protected by an **anti-virus program**. The traffic to and from the Internet is monitored. A program for network resources and traffic monitoring ought to be introduced and a server for the Internet access control ought to be established and a clear policy of staff licenses defined. On the basis of these the traffic will be filtered.

Along with protection, **permanent data availability** is another standard that the information system is expected to provide. Hardware defects and the factor of human error should never lead to the data being inaccessible for production and development. Therefore, on the *host* we have an internally developed complex application for archiving and storing data in back up copies. The same is accompanied with the **DFHSM** software (Data Facility Hierarchical Storage Management – a data archiving and storage system), which apart from the mentioned application provides for additional high-level availability of data on the *host*.

Just as regarding the issue of protection, we also need to introduce a similar system into the production on the *PC* platform. (Introducing standards, providing automatic back up for: *SQL* bases; *File* server data; user data files on workstations that require higher availability and accessibility, to a central location).

Statistical data analysis – application of specialized programs

The application of specialized programs for statistical analysis is required in order to improve the analytical performance of statisticians.

Statistical data analysis in official statistics is carried out on two levels:

- data analysis as a composite part of regular statistical activities (assuming regular periodic appearance) and
- data analysis executed upon special requests (by governmental bodies and/or other national or international users) and analysis needed for the development of existing projects or the inception of new projects.

For the first group of activities, we will use standard software for seasonal adjustments of statistical (time) series (the EU envisages the using of X-12 and TS software, in order the data from various countries may be comparable), as well as *SPSS*.

The second group is of a rather higher level and naturally requires a more developed and powerful software. The software of this kind is SAS (Statistical Analysis System), which the official statistics of Serbia used to apply from the early eighties of the last century.

It is necessary to restart the SAS software usage, especially since the official statistics of Serbia employs a sufficient number of highly qualified and experienced professionals, who are able to re-activate numerous projects, primarily those of analytical and research kind.

A minimum set of specialized SAS sub-systems for statistical purpose would include:

- *SAS/Base* – a subsystem for data handling, basic (descriptive) statistical analysis and reporting;
- *SAS/STAT* – a subsystem with all-comprising tools for statistical analysis (including those of the highest, scientific research level);
- *SAS/IML* – a subsystem for operations with matrix format data;
- *SAS/ACCESS* – a subsystem for direct access to databases (e.g. *DB2*) and/or *PC* files;
- *SAS/Warehouse Administrator* – a subsystem for handling data and meta-data in overall *ICT* environment.

The theoretical level of the *ICT* system development should include also other SAS subsystems. The only to be noted here is the following:

- *SAS/ETS* – a subsystem featuring tools for the analysis of time series and econometric surveys.

Publications processing and statistical data dissemination

In data dissemination we are expected to observe the main guidelines (for development):

- **Automating procedures.** On the level of survey planning it is required to provide that inquiries for outputs of various forms are (timely) updated, first of all to reduce ad hoc inquiries and then to respond to the inquiries (that are expected) more efficiently by standard tools and engaging several people (decentralized).
- **Standardization of output** is a prerequisite for efficient operation:
 - Standardization of texts, especially of our specific needs:
 - . Code: Unicode; windows CE/Cyrillic; etc.
 - . Multilingual options: Cyrillic – Latin – English
 - Standardized operating / supported software;
 - Selection of suitable electronic formats for preparing press and electronic materials
- **Change of prevailing media agent** for statistical data dissemination, from press to electronic. **Website**, publications on CD and the inception of new dissemination techniques (*Web* service) and new kinds of electronic publications (interactive, multimedia, etc.)
- **Introducing *print on demand***, as an effective way to meet ad hoc demands for prompt and good quality print services of small circulations, supposing new equipment is provided.
- **Archiving and easy access** to as many as possible publications (and material) in electronic form, through network. The creation of database (databases) of publications and applications for use is anticipated here.
- **Platform selection.** Primarily *PC* platform will be used in preparing outputs, except in cases when it is reasonable that the processing is directed, wholly or partly, to *host* computer.
- **“Controlled” decentralization** of material preparation for dissemination. Improved independence for divisions, according to agreed procedures, in preparing output to the extent possible (the whole output).

Project documents

Pursuant to “Standard procedures of statistical survey design and execution” we shall have improved the documents relative to the procedures of applying IT within project execution.

Therefore, special care is to be taken regarding:

- general project scheme;
- form and contents of input and the way of data introduction to system;
- databases (data files) used as consulting;
- database of survey outcome (definitions – individual data, aggregate data);
- project phases, obligations and terms;
- software to be used in certain project phases;
- list of project programs (processing segments).

In this way project documents are to be provided during the project course.

Modernization of *ICT* equipment

In order to facilitate the development of local computer network, a certain number of server computers, *PC* workstations and printers will be purchased for the Office and the regional offices as well, providing each post with a computer. The existing computers will be upgraded with certain components for data storage and memory modules. In order to provide for server reliability, units for permanent supply (*UPS*) will be purchased.

In order to connect local computer networks, communication links will be upgraded in order to meet the demands imposed on the Office, i.e. digital connections will be created or the technology of networks with package transfer will be improved. These connections require specific hardware, e.g. modem, router and Firewall, which are to be purchased.

Regarding the *host* computer, in the forthcoming period we expect the demands for: faster processing, faster access to data, larger storage space, improved print quality, variety and safety and therefore multi-processor operations will be initiated along with using adequate operating systems and software products. Also storage capacities will be extended and new printers purchased.

Staff training

Taking into account the fast development of *ICT*, as well as the fact that new and more sophisticated technologies are appearing every year, in order to keep pace with international standards and trends, we feel a need for continual education and professional promotion.

The staff training in *ICT* field will be implemented through specialized courses and individual further education, by using professional literature and the Internet. We shall carry on the practice that the employees who complete a specialized course or make individual attainment in using new software tools should transfer the basic information on the subject to other colleagues whose job is of the same or similar kind.

We shall keep on educating our employees of all statistical fields (departments) in using new information technologies.