



Economic and Social Council

Distr.
GENERAL

CES/AC.71/1999/18
23 December 1998

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 15-17 February 1999)

Topic (ii): Economic issues associated with the implementation of modern IT and management of human resources as driving forces to improve timeliness and quality of statistical data

**SOME TECHNOLOGICAL AND ECONOMICAL PROBLEMS
IN THE IMPLEMENTATION OF MODERN INFORMATION TECHNOLOGY**

Submitted by the Central Statistical Bureau of Latvia¹

I. INTRODUCTION

1. The modernization of Latvian statistical information technology (IT) was started in 1992 with the help of EC. During the first stage, the main focus was on the design and improvement of the overall IT infrastructure. The current project is the Improvement of the State statistical system (expansion/enhancement of survey activity) in accordance with "Statistical requirements compendium".

2. The Central Statistical Bureau (CSB) of Latvia holds the key role in the Latvian State statistical system. CSB is the centralized data collector from respondents and the producer and disseminator of aggregated statistical information. Main legal and methodological bases are: Law on State Statistics, Law of Protection of Personal Data, other legal acts, "Gentlemen's agreements", definitions, methodologies, classifications, lists of variables

¹ Prepared by Janis Linde and Ilmars Vanags.

(indicators).

II. THE ORGANIZATIONAL ASPECTS OF IT IMPLEMENTATION

3. CSB relations with the respondents and data users are based on varying functions, responsibilities and rights. In some cases the respondents are also users of the statistical data and vice versa. A conflict in the client interests can be seen by viewing CSB as a client system with 2 types of clients: statistical data providers and users.

4. Respondents providing information can be both a legal entity (enterprise or organization) which provides data about their activities according to the requirements of the Law on State Statistics, and a physical entity who provides information about himself or his household on his own accord or by obligation. A specific problem is the credibility of the data received from the respondents. Special methodologies and technologies are used in order to solve this problem which require time-consuming work to determine the quality of all statistical data.

5. The user of statistical data (legal or physical entity) is interested in receiving more detailed and credible data. The Law on State Statistics and other related legal acts regulate the rights and responsibilities of both parties. The main condition of CSB as a client-oriented system is to optimally meet the user interests while not increasing the respondent workload. The limitations are the allocated budget and data collection and processing technology available. In Latvia the state registers and data bases of associations and local governments are still at the beginning stage. At the same time, the request for statistical information is rapidly increasing. As most of the primary statistical data is collected using surveys, it significantly increases the respondent workload, especially for legal persons.

III. THE TECHNOLOGICAL ASPECTS OF IT IMPLEMENTATION

6. Research and analysis of work were carried out for planning the system development of CSB of Latvia. The CSB business module was designed using the CASE tools GRADE system. The relations between the object-oriented statistical sectors and the data processing functions were analysed. The Latvian statistical system can be split into 5 main domains, each of which consists of several sectors that further consist of separate surveys:

- Business statistics, breakdown according to branches;
- Price statistics;
- Social statistics;
- Population statistics;
- Macroeconomic statistics.

7. The information flows between the departments are mainly set according to the requirements of the Macroeconomics department, which is the main

internal data user and examiner. Each domain has a slightly different order of data collection and processing, according to indicators in questionnaires. The above-mentioned domains are serviced by the Statistical Business register and the Agricultural register which mainly design the frame of surveys and lists of respondents.

8. Currently, the dominant technological approaches in the CSB are not sufficiently integrated for data collection, input, control, editing and processing. They are used for individual surveys or small object groups. The individual data processing procedures have been created over a long period of time, often on different software bases, and are used by statisticians and IT specialist groups who are fully responsible for the results of their work. The primary data and aggregated data are stored in individual data files. The information necessary for the inner and external users is prepared from these files upon request. It is time-consuming work and restrains the inner and external user service. Frequent modifications of survey parameters and lack of an integrated metadata system prevents the standardization of data input and processing procedures.

9. According to the new IT strategy, the data input, storage, processing procedures and modules will be designed according to certain standards. The standardized statistical data production process at CSB will consist of the following main procedures:

- Survey design and preparation (centralized);
- Maintenance of the registers and respondents selection (centralized);
- Data entry and initial checking (decentralized);
- Data validation and editing (decentralized and centralized);
- Statistical data storage and processing (centralized);
- Statistical data analyses (centralized);
- Data dissemination to external users (centralized).

10. In 1998, 154 statistical questionnaire forms were used for the collection of primary statistical data. These were sent to the respondents, mainly to legal persons, and the responses were received by mail (mail out-mail in data collection method). More than 20 thousand respondents were interviewed during one year (not including the register surveys). Part of the respondent forms, mainly from physical persons, were filled in with the help of interviewers. During one year CSB of Latvia collected 151.6 million indicator values for 29,000 statistical units. 24% of these were received electronically, the rest of the data was entered manually. The large volume of raw data, the need to carefully control and check for errors and to re-contact respondents necessitates the creation of a decentralized data input and control system. The main objective of the system is to choose an optimal distance between the data source and the place of data registration.

11. Using a decentralized data input system, it would be much easier to contact the respondents, offer them methodological consultations and clarify

the errors. The non-response and respondent workload would therefore be reduced. The current practice shows that the number of errors can be very high. The clarification and correction of errors is very time-consuming, especially in financial reports where the indicators are mutually linked. The mentioned activities can only be carried out by qualified statisticians using PCs with special data input programmes, which simultaneously check the data, named Computer-Aided Data Input (CADI). The statistician is usually specialized in a specific statistical sphere, works with a known group of respondents and does the methodological work with the respondents. The operator's work, which focuses on rapid input, is not sufficiently effective in such cases.

12. The decentralized data input system includes 28 regional statistical offices with 5-9 staff working in each administrative region. Their main task is to collect raw data from the respondents, to control the data and enter it into PCs. Such a decentralized statistical data collection system has proven to be efficient over a long period of time. From the governmental viewpoint, it is positive to create new work places in the regions where the employment level is considerably lower than average.

13. The design and maintenance of a modern IT infrastructure based on the latest developments in telecommunications is very complicated: it relies on implementation of the state information programme and requires additional investments. The latter are also necessary for the IT training of the regional staff. The central office will reduce the volume of input data. A data administration system needs to be created in order to manage and co-ordinate the data flow from the centre to the regions (applications) and back (data files). Currently, on-line communication with the regional offices is not very efficient due to the low data exchange speed. Therefore, data is exchanged in the form of files utilizing the data base replication mechanism.

14. The development methodology of the CSB Information System is based on the following stages:

- i) User Requirements Specification: information system is developed to meet specific user needs.
- ii) Functional Requirements Specifications: this is a formal technical document, which reflects the user requirements expressed in stage i).
- iii) System Architecture and Design: this is an overall design of the system, derived from the Requirements Specification. The architectural design specifies the IT infrastructure and the main modules of the information system.
- iv) Module Design: the main elements of the detailed design are the Data Flow Diagram and the Data Model (like ER model). In this stage the software modules are identified and their functions specified in detail. The modules will be further broken down into Basic Units (sub-modules) suitable for development.
- v) Coding of Units: the basic units defined in stage iv) will be developed.
- vi) Coding and Testing of Units: the units coded in stage v) will be tested against their formal specification.

vii) Integration: the units will be compiled into modules, and the modules brought together to form the information system. The interface between components must be developed and tested. The validity of the system will be checked against the original architectural design and modified as necessary.

viii) Testing: the full system will be tested to ensure that it is in accordance with the Requirements Specifications.

ix) System Acceptance: the final version of the system must be approved by users. To facilitate this, the users should be assisted in preparing a plan for an acceptance test of the system. At the conclusion of this stage, the information system enters into production.

15. In practice, the flow between these stages very often is an iterative process. For example, consideration of architectural design and the constraints imposed by the available IT infrastructure may necessitate changes to requirements specifications. These will in turn necessitate changes to user requirements, which should be agreed upon before development continues. In the CSB IT project, the third stage has been completed. Functional requirements have been prepared and the main modules have been defined together with their functions and interfaces.

16. The data flow analysis and CSB business module analysis shows that there is a necessity to store, process and analyse data using data warehouse technology. Several object-oriented operational databases for the storage of primary data need to be created, as well as a central aggregated database and a special database for external users. A metadata base is necessary in the implementation of such information technology. The main tasks for the CSB staff for the current period are as follows:

- To develop an integrated classification system;
- To develop an integrated system of statistical indicators responding to the Eurostat requirements. To define the relationship between indicators;
- To harmonise rules for data aggregation and to define the most frequently used formulas and methodology

17. The implementation of the above-mentioned activities will create a basis for the integrated metadata base. Its development will enable a standardization of statistical data processing and storage technology.

IV. THE ESTIMATION OF IT EFFICIENCY

18. The implementation of modern IT in a statistical office is not an end in itself, but a means to achieve effectiveness and efficiency. The effectiveness is the degree to which the predetermined goal of the state statistical system is achieved, that is, to supply domestic users and international organizations with timely, correct, complete, unbiased and easily comprehensible statistical information in order to make decisions on the development of the national economy and social sphere. Efficiency is measured by the results obtained in relation to the resources used. Different criteria, methods of calculation of

indicators of efficiency, benefits and costs, and necessary conditions should be used depending on the purposes of calculating the economic efficiency of IT. The choice of an economically efficient variant of IT in a statistical office, and the estimation of actual or planned efficiency of automation should be based on a cost-benefit approach. Listed below are the main categories of efficiency of IT in a statistical office:

- Efficiency from the standpoint of the state as a whole and efficiency from the standpoint of the statistical office (or other organization);
- Efficiency of a project (economic measure, technical solution) and efficiency of economic-financial activity of the statistical office;
- Efficiency for selecting an alternative solution, planned efficiency and actual efficiency;
- Integrated effect, yearly effect and average yearly effect;
- Absolute (common) efficiency and relative (comparative) efficiency;
- Efficiency of statistical information system, subsystem, survey, combination of tasks, task and technological operation;
- Efficiency of a computing unit, hardware and software, data base, system of communication, etc.

19. Taking into account the fact that statistical information is a public good, the implementation of modern information technology in statistical offices should be evaluated firstly from the standpoint of the state as a whole (economic analyses of project) and secondly from the standpoint of the statistical office (financial analyses of project, cost recovery analyses). In the first case, costs and benefits should be accounted for not only in the statistical offices (central and regional), but also in relation to respondents, users of statistical information, administrative registers and other participants. It should be mentioned that Council Regulation (EC, Euratom No 58/97 of 20 December 1996 concerning structural business statistics) states in paragraph 2, article 7 that quality evaluation shall be carried out comparing the benefits of the data availability with the costs of data collection and the response burden on businesses, especially on small enterprises.

20. Bearing in mind the limits of capital investment, particularly in the transition countries, the criteria of the efficiency of implementation of modern IT could be:

$(B-C)/K$ -> maximum,

where: B - discounted benefits over the life time of the project;

C - discounted current (periodical) costs over the life time of the project;

K - discounted capital investment.

21. Many statistical offices (including CSB of Latvia) are chiefly financed from the state budget and disseminate statistical information to government institutions free of charge. In most cases, it is very difficult to determine the benefits of the use of statistical data. Therefore possible calculation of

efficiency could be the following:

$$DC/Ka = (Cb - Cc)/(Kc - Kb),$$

where: DC - savings of average annual current (periodical) costs;

Ka - additional capital investment;

Cb , Cc - average annual current (periodical) costs in base and comparable variants;

Kc, Kb - capital investment in comparable and base variants.

22. To calculate the above-mentioned indicator (rate of return on additional investment) correctly, the same input (primary data) and the same output (aggregated data) should be used in both the base and comparable variants. The calculation methods differ in the case of the efficiency estimation of an information system and in the case of separate tasks, surveys or subsystems.

23. The recommended concepts for the calculation of efficiency are the following:

- Direct effect (in the sphere of data processing) and indirect effect (as a result of using higher quality information, mainly in the sphere of production and services);
- Past (historical) costs and future costs; fixed costs and variable costs;
- Pre-processing costs (training of personnel, upgrading the software etc.) - a characteristic feature of information systems;
- Norms of costs relating to time unit of work (machine-hour, norm-hour, man-hour);
- Standard forms of calculation of economic efficiency of data processing.

24. The most cost-efficient variant of the implementation of modern IT in the statistical system should be chosen taking into account only relevant costs, that is, expected future costs (which differ among alternative variants). The past (historical) costs should be used to calculate actual efficiency; they may be an important data source for predicting future costs.