

**Distr.  
GENERAL**

**CES/AC.61/2001/27  
13 August 2001**

**Original: ENGLISH**

**STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR  
EUROPE**

**CONFERENCE OF EUROPEAN  
STATISTICIANS**

**Joint ECE/EUROSTAT/FAO/OECD Meeting  
on Food and Agricultural Statistics in Europe  
(Geneva, 17-19 October 2001)**

**COMMISSION OF THE EUROPEAN  
COMMUNITIES (EUROSTAT)**

**FOOD AND AGRICULTURAL  
ORGANISATION (FAO)**

**ORGANISATION FOR ECONOMIC  
CO-OPERATION AND DEVELOPMENT  
(OECD)**

## **IMPACT OF HIGH TECHNOLOGY ON USE OF EXISTING DATA SOURCES**

**Invited paper submitted by Finland\***

**Summary:** Information and Communication Technology makes it possible for us to use existing data sources in statistics. This new statistical process requires a legislative framework, cooperation and appropriate information systems. The setting up and maintenance of registers calls for good planning that takes the potential of joint use into account. Cost savings do not become apparent until later, because exploitation of the new technology calls for development work and costly investments in software and hardware.

Statisticians whose work involves the use of existing data require familiarity with a multiple range of fields encompassing agriculture, statistics and IT. The recruitment of such widely qualified persons is a challenge that our statistical organisations will have to face in the near future.

---

\* Prepared by Mr. Esa Ikäheimo, Information Centre of the Ministry of Agriculture and Forestry, Finland.

The quality of statistics will become an increasingly prominent issue, as the use of existing data sources distances the compilers of statistics from the original data, making it increasingly difficult to check their quality. The users of statistics require information ever more rapidly, and thus the time available for quality checking is cut to a minimum.

High technology facilitates data collection and ensures a better service for the users of statistics because the information can be extracted rapidly from existing databases and made available immediately via the Internet.

In the near future it will also be possible to make statistical returns via the Internet. Interaction between the compiler and the user of statistics will become closer, which will increase the motivation to respond to statistical surveys.

## **I. Introduction**

1. High technology with its various applications is becoming increasingly pervasive in business life, and this includes the production and dissemination of statistics. I intend here to look at the effect the use of high technology in the production of statistics— chiefly Information and Communication Technology (ICT) – has had on statistical work. In point of fact, electronic data transfer is a basic requirement for use of our current databases. It serves the production of statistics in many ways, notably in the use of existing data sources, which has become increasingly common in recent years

2. High tech applications in agricultural and land use statistics include satellite technology and the approaches used in Area Frame statistical work such as GPS. A new challenge that the compiler of statistics faces almost daily is the e-Society, or e-Economy. I cannot imagine how we would describe this feature in statistical terms without the use of ICT. Electronic data transfer is currently widely used to disseminate statistics via the Internet or e-mail, for instance. The utilisation of existing data with the aid of ICT has now superseded traditional methods in the production of statistical publications.

3. New opportunities for the use of existing data in agricultural statistics opened up with the implementation in the early 1990s of the EU's Common Agricultural Policy (CAP), and the compiling of the information on individual farms into computer-based data warehouses. Each Member State was obliged under the relevant EU legal act to set up an Integrated Administration and Control System (IACS), in which the data were recorded in electronic form. As the system included not only general facts about each farm but also a large amount of information about the farm's production resources (use of arable land, livestock, etc), this information began to find use in statistics, too.

## **II. Preconditions**

4. The flexible use of existing data sources places demands on the legislation governing registers and their administration. One of the biggest obstacles to joint use was the provision restricting the use of registers to the needs of the keeper of the register.

5. With the spread of electronic data transfer, joint use became technically increasingly feasible, and a start was made on removing the legal obstacles in its path. Finland's new Statistics Act empowered statistical bodies to utilise the data held by authorities should it be necessary for the compiling of statistics and to help avoid duplication of data collection. Under Finland's statistical legislation, the use of an existing data source is not simply an opportunity for statistical bodies but a binding obligation.

6. The basic precondition for the linking of registers – irrespective of the technology used for that purpose – is the unambiguous identification of the registered units. In registers of people the identification is done using the personal identity number. This code is unique and defines the person exactly. The use of codes for businesses is not always so easy as different keepers of registers, such as statistical bodies, tax authorities or the National Board of Survey, define a registered unit, e.g. a farm, differently from each other. Indeed the need to harmonise the definitions and classifications used for registered units has become increasingly obvious in recent years. To this end, we have set up a project in Finland to integrate the farm register with the general register information system, which includes all businesses. In terms of rural policy an integrated business register system is important because it will enable us to consider businesses operating in rural areas as entities of a single owner or holder and to look at the effect the measures applied have on them.

7. From the technical point of view it is essential that the data warehouses should be appropriate for joint use, that the necessary hardware and software should be available, and that personnel should be sufficiently competent in the use of IT. Close cooperation between the producer of the information and the compiler of the statistics is a basic requirement. The need to link registers should be taken into account right from the planning stage, for instance, in the definition and classification of registered units and in the codes used.

### **III. Risks and opportunities**

8. Information technology and its vast potential improve statistical production in many ways, e.g. by reducing duplicate data collection. That is not to say, however, that we get the information we want simply by 'pressing a button'. The use of existing registers and other data sources has made statistical work a complex process in which the compiler of the statistics has to be familiar with the definitions and classifications of the units in various registers and also with the technological approaches used for them. The risks involved in data transfer and information security are increasing and call for special arrangements.

9. When we use existing data sources, we often assemble the information on statistical units from different registers. A problem in joint use may turn out to be the quality of the data sources, as this cannot be checked in the same way as that of material we collect ourselves. Yet another problem is that the coverage and periods of reference of different registers may vary. Naturally, the completion of statistics depends on the last register completed.

10. When the statistical bodies themselves collect information, for instance, by postal or telephone enquiry, it is easier for them to check the original data and so keep the other elements of the statistical process under control.

11. The expansion of ICT has not only provided new opportunities for the use of existing data; it has completely altered the whole basis of statistical production. Our traditional statistical system has reached a turning point and we should perhaps be asking whether we should reject the old statistical surveys and registers altogether and coordinate existing data into statistics even though the sources of the data do not satisfy the requirements of homogeneity. The demand seems to be for statistical information produced within very short intervals, even at the cost of quality. This refers in particular to the statistical products, e.g. market prices, of the e-Economy area.

12. It is usually in the interest of those providing the information to fulfil their statistical obligations with as little inconvenience as possible, from their own information system direct. Today's data transfer technology makes this possible. The statistical production process should indeed be developed as a whole, starting with the information flows and warehouses of the provider of the information and proceeding to the user of the statistics via the editing and tabulation done by the statistical body. The motivation of the providers of the information could be increased by giving them a greater say in the statistical process.

13. Upgrading the information systems of statistical bodies to enable them to handle data transfer often requires the whole database to be restructured. This in turn calls for considerable input of human effort and the acquisition of hardware and software. Further, there must be capacity to serve those who need the statistics. They expect readable analysed information on line that they can then rework for their own needs.

#### **IV. Impact on personnel strategy**

14. The increase in the use of existing data has eliminated the mailing of statistical forms and their processing and storage as data files. For example, 95% of the increase in the data resources of Statistics Finland comes from other registers; only about 5% is collected by statistical surveys. In their personnel structure, statistical offices naturally have to keep up with changes by increasing their intake of IT experts and cutting jobs involving the pre-processing of statistical information.

15. Experts in agricultural statistics are expected to be familiar with both agriculture and statistics. The joint use of registers calls for IT skills, but demand for such persons has increased very rapidly in other fields, too. Training has not kept pace with changes in the labour market, and as a result we are now faced with a situation in which demand exceeds supply. One outcome of this is that the salaries of ICT experts in the private sector have shot up. The public sector and statistical organisations have to take account of these changes in their personnel strategies. It is, however, difficult to devise a remunerative system that would be equitable for all members of staff while also satisfying the special needs caused by the high market salaries received by ICT experts.

#### **V. Impacts on cost of statistical work**

16. The use of existing data in statistics calls for new solutions in both IT and data transfer. As the rural business register was gradually introduced ever more widely in Finland, we found it necessary to maintain two systems during the transitional period to secure the quality of the statistics acquired with the new method. This naturally increased the cost of compiling the statistics.

17. The acquisition of information from existing data sources reduces the direct costs of collecting and processing information. Extra costs arise, however, when we have to develop new information systems and purchase the equipment and programs required by those systems. The use of high technology calls increasingly for the recruitment of staff with special skills.

18. The IT field is advancing rapidly, so much so that even information and software acquired this year have to be updated at regular intervals. Likewise staff employed in IT need more training than others in order to stay abreast of developments in their field. Such training is expensive and raises the cost of using existing data sources.

19. Once the systems required for utilising existing data sources have been developed, we can expect costs to come down permanently, providing, that is, that the sources can be utilised free of charge. At the outset, the benefits will be seen most clearly in improved services.

-----