



Economic and Social Council

Distr.
GENERAL

CES/AC.71/1999/8
1 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 (i): The impact of Internet on the statistical production and dissemination process

THE IMPACT OF THE INTERNET ON THE STATISTICAL PRODUCTION AND DATA DISSEMINATION PROCESS

Submitted by Hungarian Central Statistical Office¹

1. The Hungarian Central Statistical Office (HCSO) has been publishing its data on the Internet for two years. The address of our web server is: <http://www.ksh.hu>. We belong to those European statistical offices which early on recognized the possibilities of the fast forwarding of large quantities of data to users through the Internet. The Stadat system installed by the HCSO is in continuous development both in terms of data contents and tools.

I. A BRIEF INTRODUCTION TO THE STADAT SYSTEM

2. The Stadat system is static, i.e. it consists of ready-made tables prepared by the different departments of the HCSO which can be accessed via the local computer network. The tables are produced in the database by linking tables covering various periods (months, quarters) so that they can be retrieved as a single table. The database also includes tables containing time series. There it is not necessary to select any period because data on the

¹ Prepared by Zoltán Papp and Gyongyi Takács.

various years, quarters and months can be found within one table.

3. The system exists in English and Hungarian. Part of the data is accessible free-of-charge, the other part against payment. The free-of-charge data include general information, the main figures introducing Hungary, the latest data of interest to the general public, and the most important annual economic indicators. The chargeable part includes the following:

- ON-LINE access to monthly, quarterly and annual data in the form of tables dating back to 1996;

- download of first releases. First releases recorded in Winword and Excel formats encompass the latest statistical data of the HCSO. The publication consists of an analytical section (about 5 - 10 tables) and a methodological section. The latest publications (around 30) can be downloaded in Hungarian, some of them are also available in English (about 6 can be downloaded).

4. The HCSO is preparing the English versions of the first releases (issued only in Hungarian as yet) and so the English version of the system is continuously being enlarged. As the data of the HCSO are used, via Internet, by an increasing number of people, comments on and demand for the data are steadily growing. Further details of this will be given later.

II. DATA SOURCES OF THE STADAT SYSTEM

5. As mentioned in the introduction, the Statat system is static, which can be seen from the methods used for processing, storage and publication of data. The tables in the ON-LINE section of the Statat system are taken from the publication in which the data in question have been first published. This is important because the HCSO publishes the same data in several publications.

6. The monthly data are published in the first releases, and then in the "KSH Reports" and, finally, in the "Monthly Bulletin of Statistics". The first releases are prepared and recorded in Winword and Excel format by the professional departments of the HCSO.

7. After the first release is approved by the President or the Deputy President of the HCSO, it is made available by the authors on the HCSO local network. The other HCSO experts can download the release from the Statat system to their own PCs, and on the Internet server using file transfer (FTP). External users can download the first releases from the Internet.

III. TERMS OF PAYMENT OF THE CHARGEABLE BLOCK OF THE STADAT SYSTEM

Prices, Allowances

8. The use of the ON-LINE access to monthly, quarterly and annual data costs USD 80.- per month. As there is no accounting system connected to the network, we are charging the flat rate. In this way, subscription to the system is favorable mainly for regular users. Those who only use the system occasionally can subscribe for a single month, but they must indicate their intention well in advance. HCSO grants a 60 per cent price allowance to the universities and colleges, and 50 per cent to the central administrative bodies.

Contracts

9. HCSO concludes a contract with the users of the system which can be terminated without any explanation at the end of a quarter.

Identifiers

10. Access to the system is possible through a pair of identification codes which are coupled with the various services and which are mentioned in the contract.

Entry Forms

11. The chargeable block of the Stadat system contains the Order Form which can be printed and sent after completion to the HCSO by fax or e-mail.

IV. USERS OF THE STADAT SYSTEM

12. The majority of users of the system belong to the entrepreneurial sphere: 55 per cent of a total of over 100 subscribers, 20 per cent are educational institutes and 16 per cent the state administration.

13. It appears from statistics that the entrepreneurial sphere badly needs rapid, precise and easily accessible data. Given their scope of activity, the simplest way for them to access statistical data is to receive the data through the Internet. The HCSO disseminates the data in paper or fax form to the state administration and to the editors of newspapers more-or-less free of charge. Therefore, there are no subscribers among the editors of newspapers. However, the HCSO does not forward information to enterprises, so their only access to our data is through purchasing it in some way. As it is easier to subscribe to an Internet database than to buy and store books and publications, they choose the Internet. It is much simpler to look for retrospective data in a database than in a mass of books.

14. There is keen interest among foreign users as well. International organizations such as the UN and the OECD receive the HCSO's data through the

Stadat system. We have learnt through feedback channels that they are satisfied with the services of the system.

15. Our foreign subscribers enlighten us on the faults and shortcomings of the system. We know from questions asked that HCSO prepares its statistics differently from West European countries, e.g. in Hungary GDP is calculated on an annual basis while our partners require GDP data on a quarterly basis. In Hungary the indices are calculated based on the corresponding period of the previous year and, in some cases, with that of 1990 or 1995. But it is necessary to calculate indices for several basic years.

16. The questions received by the Statdat Office on the Internet are forwarded to our Dissemination Department and to the responsible departments of the HCSO. These indications are important for us because of our EU-association.

V. COMPARISON OF INTRA-INSTITUTIONAL COSTS OF PAPER BASED AND ELECTRONIC PUBLICATIONS

17. It is important that users can rapidly access statistical data, but another issue of an equal importance is the question of financial matters. First releases of the HCSO can be downloaded from the Internet at any time. This is important because only the latest publication is available in downloadable form in the database. As the first releases are published monthly, they are changed every month. The dates are listed in a release calendar on the Home Page of the News Block. Every user can find out on which date or, in the case of consumer price indices, at which hour the data he/she is interested in becomes available.

18. The price of a first release is HUF 200.-. If any user wishes to buy all 36 publications downloadable from the system, the business price is HUF 10,800.- including only one copy of each publication. In an agency where staff members have work places far from each other, several copies of the same publication must be bought. If an agency subscribes to the Statdat system, any responsible staff member can download as many copies as required of all the publications at a price of HUF 5,000.-.

19. It is essential that, using a single user name-password, several staff members can simultaneously have access to the system. Not only a single person but a whole institution is thus authorized to use the same user name-password. It is obvious that costs can be saved through the use of Internet.

20. Naturally, it is advantageous to the persons using the data not to have to search for the required publications, and all information is available to them at any moment. Last but not least, large quantities of paper can be saved.

21. It is, of course, in the HCSO's own interest to have an ever-growing number of users connected to its Statdat system because it is also a source of income. The Ministry of Finance stipulates an income for the HCSO as equal to

10 per cent of the total financing of the HCSO from the state budget. Naturally, the data of the HCSO are available at the Library of the HCSO free of charge. The HCSO's Library has 40,000 visitors annually.

VI. MODIFICATIONS AND FURTHER DEVELOPMENT OF THE STADAT SYSTEM

22. On the basis of the experience gained during the two years of the Statat system's operation, we are continually modifying the system. Our main objective is to depart from the "publication approach". The users need to have access to the latest monthly data as soon as possible and to receive time series from the annual data as early as possible. In order to produce long time series in tables the data must be homogenized. When the annual data were published in yearbooks, the problem of comparability did not exist. But if, on the Internet, the data appear in several tables, they cannot be copied together into a single table because the data have been prepared using different methodologies and over different periods of time.

23. We have decided to indicate the homogenized and modified data which have been included in a table of long time series with a different letters in order to distinguish them from other data.

24. As mentioned earlier, the system is static, i.e. the database contains ready-made tables. This has some advantages as many users need the data collected in one table. They could compile the table by retrieving data from a dynamic database, but that would require typing a series of questions which takes time and leaves possibilities for errors.

25. While taking into consideration these demands, we have to depart from the "publication approach" and must create special tables for electronic dissemination. As there are no limits to the size of a table on the Internet, long time series can be released in a single table. For the long tables, we have solved the problem of freezing the table headings and rolling the data under it. We were assisted in this work by the Statistical Office of Stuttgart, Baden-Württemberg.

26. The users are demanding more and more data on regions and provinces, for both individual regions/provinces and comparable data for all the regions.

27. The HCSO is further developing its database system in order to establish a database for dissemination purposes along with the database for the production of statistical data. Work has been started in two directions:

- principles of the contents and functions have been determined;
- survey of the technical facilities of the implementation and the examination of potential software has been carried out.

28. The principles can be summed up as follows:

- The database for dissemination covers all the important statistical

topics independent of the data production sphere and the technical environment of data processing. The database will be prepared and systematized according to users' requirements. The contents of the data are described by metadata. The main elements of the database for dissemination are the following:

- the part intended for external users: narrow data contents, high aggregation, does not include confidential data, bilingual, available through the Internet;
- the part intended for internal users: it has to allow analysis of micro-data; it has to provide a possibility for flexible and fast retrieval/analysis of the most important aggregates; micro-data should be available while analyzing the aggregates.

This scope of problems is related to the term "data warehousing".

VII. DATA PROTECTION

29. The HCSO has implemented a two-level protection system for the local area network. The first level works in the Information Center of the Prime Minister's Office. This Firewall is very complex and filters attacks from the global Internet zone. In the HCSO, every host is protected from the Internet zone except the WWW. The second security level, developed by HCSO, is a filter against the network of Prime Minister's Office.

30. The illustration below shows the computer system of the HCSO:

