



**Economic and Social  
Council**

Distr.  
GENERAL

CES/AC.71/1999/2  
4 February 1999

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 STATISTICAL ORGANIZATIONS**

Submitted by the Federal Statistical Office of Germany<sup>1</sup>

**INVITED PAPER**

**I. THE INTERNET AS A TOOL FOR STATISTICAL OFFICES**

1. Accessing the desired information just by pressing a button - this is a vision that became reality through the Internet. The Internet helps to channel the growing flood of information and to make acquired knowledge available for further use; moreover, it supports the process of searching for existing information: The World Wide Web as a universal and particularly user-friendly communication service of the Internet makes networked computers accessible to a large group of users by "mouse click". This was a privilege for a rather small number of experts still some years ago. For users who are interested in the information itself rather than technical details of information supply,

---

<sup>1</sup> Prepared by Mr. Sarreither and Mrs. Novak. The document is an update to the paper CES/1997/7 that was considered at the 45th Plenary Session of the CES, Geneva, 10-12 June 1997.

the World Wide Web is a suitable platform because of its easy operation .

2. Numerous providers have made it their aim to fill the technical infrastructure with adequate contents so that the Internet can successfully be employed as an information tool. No institution disseminating information will be able to ignore the electronic data highway. Almost all of the statistical offices, too, have accepted that challenge, joining the network with their own programmes. First, they use this channel to disseminate information, employing the Internet as a medium of public relations. Second, the Internet is suitable for commercial marketing because, similar to printed or other electronic publications, it also offers the opportunity to supply data against payment (electronic commerce).

3. Since March 1996, the Federal Statistical Office has been present in the Internet as a data provider with its own server. The Office has acquired knowledge of how to use this medium, so that it is possible now to exchange experience with other institutions. What can be said already now is that the expectations regarding the acceptance of the German statistical data supply in the Internet have even been exceeded. The Internet offers a suitable tool for marketing, sale, support and data gathering.

## II. REQUIREMENTS TO BE MET BY STATISTICAL DATA OFFERED IN THE INTERNET

### II.1. Requirements in terms of contents

4. The World Wide Web service provided by the Internet is particularly well suited as a medium for disseminating statistical information, as it permits supplying texts, tables, charts and other multimedia objects in an up-to-date and low-cost manner to any number of recipients all over the world at the same time; moreover, it allows establishing connections with other providers. Users may access the desired information in a time-saving manner. For the medium to be accepted it is necessary that, in addition to quick and user-friendly access, and beyond the usual publishing rules, some technical and user-specific factors are taken into account and that the information is processed in a specific way to suit this medium. This includes:

- **A clear user guidance:** The contents have to be clearly structured and limited to the essential information. They can be presented in "small portions" only. What has proved useful is a hierarchical structure guiding the users according to their objectives. The use of hyperlinks, modern styled pages using dynamic HTML and other navigational features gives the user the complete control of knowledge discovery.
- **Using search devices:** For statistical information in particular, the use of "search engines" is ideal, because users are guided directly to the desired texts or charts after entering a key word. The statistical offices can improve this way of information retrieval in defining harmonised and consistent metadata all over the broad area of

statistical surveys.

- **Taking account of the worldwide structure of the group of users:** An information tool available all over the world has to take account of the different situations of potential users. This ranges from language to differences in background knowledge about national particularities. To allow interpretation of the statistical data, not only the data themselves but also data descriptions are necessary. Explanatory notes and documentation should be accessible in the background. This will be explained in greater detail below .
- **Quick operation is more important than color images:** Not every Internet user has sufficiently quick access to the data network. For users with rather slow access, it is hardly reasonable to request pages with a lot of graphics. This is why the information offered should not be "over-ornate in terms of graphics". It must be possible for users to find their way through the program largely without the graphical component offered as they should have the opportunity to have the pages represented without graphics.
- **Consistency of the stock of information:** A growing information supply easily develops a trend towards redundancy and inconsistency. The entire stock of information must be monitored and needs permanent maintenance, so that uncontrolled developments are avoided.
- **Timeliness:** It is necessary to always keep the data supply up to date, in order to prevent users from ignoring the data supply in disappointment after initial enthusiasm. We have to guarantee that our information production process integrate the new medium. There are strong efforts in defining a source document type which can be translated very quickly in various media types. We have to mention the opportunities which is given by the new XML standard.

## II.2. Requirements in terms of data quality and comparability

5. It is not only since the rapid growth of the Internet that the countries of the world have come closer to each other; a crucial factor for this development have been the real trade and financial flows. As a result, the demand for up-to-date and comparable high-quality data is growing. Examples are the data required to judge whether a country meets the convergence criteria of the Maastricht Treaty and the requirements of the International Monetary Fund regarding the supply of business indicators.

6. The Internet provides the technical platform for rapidly supplying the decision-makers in politics and business all over the world with the data they need. This makes it even more important that standards of disseminating statistical data are met; such standards are demanded, for instance, by the International Monetary Fund.

7. A particularly important element in this context are metadata. What is indispensable is especially information on data quality. This includes information on breaks occurring in time series, sampling errors, cut-off limits, etc. Without easy access to this additional information, the mere data are useless.

### **II.3. Data safety**

8. The networking of many private nets to form a worldwide network on the one hand opens up enormous opportunities of serious utilisation while, on the other hand, it creates problems of data protection and data safety. These aspects are important also for statistical office's employing the Internet as a medium of publication.

9. Two sets of problems should be distinguished here:

- Using the Internet as a transportation network: Since the Internet is an open and decentralised network without any central network operator, it is generally impossible to clearly identify Internet users. Any user may read and modify the message of another user which passes a router accessible to him. There are sufficiently safe methods to overcome these shortcomings; standardising them is a precondition for the commercial utilisation of the Internet. This set of problems becomes relevant where the Internet is used as a tool for transmitting confidential data or where data are supplied against online payment. Standards like SET (secure electronic transaction), digital signature procedures and various kinds of "secure servers" have to be implement in the offices.
- Protecting an institution's internal computers against unauthorised access from the Internet: Any use of Internet services by an institution or the operation of a computer providing information requires a connection between the computer and the Internet. If, at the same time, that computer is connected to the institution's internal network, there is also a connection between this network and the Internet. This connection is not only accessible to authorised users but it might also be used for unauthorised access.

10. Since, for operational reasons, this connection, cannot be interrupted, protection can be safeguarded only by strict control and security measures. For this purpose, clearly defined and controlled "firewalls" have been established. These computers control the data flow, refuse unauthorised access, permit authorised access only and detect intrusion.

## **III. DATA OFFERED BY THE FEDERAL STATISTICAL OFFICE IN THE INTERNET**

### **III.1. The global aim**

11. The aim of the Internet activities of the Federal Statistical Office is the worldwide presentation of a highly up-to-date information supply that has

been designed in a user-specific way and is extended step by step. The final product will be a comprehensive-and-user-friendly information system whose data can be accessed all over the world mainly in the form of a statistical database.

### **III.2. Technical implementation**

12. The server of the Federal Statistical Office can be accessed through a connection to the Deutsches Forschungsnetz (German research network) in the World Wide Web service provided in the Internet. Before the Internet connection was established, a specific firewall computer had been installed. The Internet computer providing the generally accessible data supply is located before the firewall computer. It does not contain any data to be protected. For the time series service explained in detail below, an additional World Wide Web secure server has been installed.

13. The information supplied by the Federal Statistical Office in the Internet is compiled at workstations. The product is transferred to the internal computer before it is mirrored to the external computer protected by the firewall computer.

### **III.3. The homepage**

14. The programme parts currently available are indicated on the homepage. The homepage is the ,Federal Statistical Office's visiting card in the World Wide Web. It contains a symbol (logo) used in the entire Internet data supply, which clearly indicates the Federal Statistical Office as the information provider.

15. First of all, the user can choose between the German and English language version of the information supply. While structured access to the information supply is enabled through "mouse click" on the desired menu item, it is also possible to enter one or more key words, start an inquiry for all text elements through a "search machine" and thus get direct access to the information pages. What can be accessed are statistical data and "service pages" providing, among other things, the background knowledge required.

### **III.4. Service pages**

16. Under the key word "About us", the legally defined function of official statistics and the division of labour between the Federal Statistical Office and the statistical offices of the Lander within the federal system of the Federal Republic of Germany are explained. In addition, the various stages of producing a federal statistics are presented and issues of compulsory response and confidentiality of statistical data are discussed. This programme part is particularly important for users who are not acquainted with the national particularities of German official statistics.

17. Under the term "Helpline", users have the opportunity to directly

contact competent staff members at the Office; this is an improvement of the current practice of putting through phone callers to competent personnel. A decentralised structure of electronic communication (via "E-Mail") through distributed access to the Internet is established over the local area network of the office.

18. In the sections "Bookstore, Mediastore", important statistical publications in printed and electronic form are presented. It is planned to integrate the entire list of Publications and to add an order service.

19. Under "Events" the participation of the Federal Statistical Office in specific meetings and fairs is indicated, so that all those interested in statistics have an additional opportunity of making contacts in the statistical sphere. Information on important national and international trends in the field of official statistics and on the activities of statistical bodies, as well as reviews of statistical-methodological literature are offered under the key word "Science".

20. Due to the decentralised structured of the Internet, links with other data providers, in addition to the commercial "search machines", are critical tools for the search for information in the Internet. The menu item "Statlink" provides the opportunity to change direct to the information programmes of other statistical offices and selected institutions in Germany and abroad. Access to the statistical offices of the Lander is enabled not only through links within the texts, but also through an image map.

### **III.5. Statistical results**

21. The "Hot News" provide the media and the general public every day with the latest important results of official statistics. This presentation of data involves no delays and is made possible by a decentralised connection of the press office to the internal network. The presentation of all press services from subject-related and chronological aspects is supplemented by a weekly schedule of publications; this permits early planning of requests to be made at the time of data publication. Important press releases are enriched with additional tables containing monthly and annual data, particularly to show economic and price-statistical trends.

22. Under the key word "Basics", key data and trends of demographic and economic developments are presented in a clear and easily intelligible form by means of texts, tables and graphics based on annual data from all important areas of official statistics. Hyperlinks provide the opportunity to go directly to another position of the same information page to another document (e.,g. an explanatory table) or to the information supply of another provider (e.g. a Land statistical office). This technology is used for the basic data to offer the user a choice of different presentations of subject-related contexts, i.e. as texts only with supplementary tables (if applicable) or with illustrative graphics; such charts will be more widely used in the future, although they require more time for data transmission. What is also being set

up are supplementary methodological notes, e.g. on the index of consumer prices. They are designed to answer frequent questions on the item concerned, thus relieving the conventional information services operating by phone or mail.

23. Since October 1996 it has been possible to access important economic indicators under the menu item "Indicators". The time series have been selected for specific target groups and have been designed to meet the basic requirements of business cycle research, banking, the economic sectors and the media. What will be taken into account more specifically in the future are the requirements of the International Monetary Fund. Unlike the statistical basic data, the tables which contain economic indicators and have to be updated every workday are provided in an automated process from the time series of the statistical database STATIS-BUND.

24. Since mid-March 1997, Internet users have had the opportunity under the menu item "Time series" to access over 1 million time series of the STATIS-BUND database covering all areas of official statistics. Economic data are the core of this range of information. Long-term annual series, part of which date back to 1950, and short-term monthly or quarterly series provide the basis for manifold studies over time and analyses. The time series service is supplemented by a catalogue of definitions explaining the variables and terms used in the tables.

25. Unlike the other parts of the Internet programme, all of which are provided free of charge, the new time series service comprises a free documentation and research system and a data ordering and collecting service by means of file transfer against payment. To access this information, users load the desired data packages onto their own PC and convert the data by means of a utility which is provided too.

26. Data packages can be ordered only by users who have previously concluded a user contract with the Federal Statistical Office. The contract form may be requested online from the Internet pages of the time series service. As an access authorization, the user will then receive a personal user ID and a password from the Federal Statistical Office. The price structure for this service is an annual basic price plus a variable price component depending on the actual use. Payment is made in the conventional way. New forms of encashment for example micropayments have to be investigated for their usefulness in the statistical offices.

#### **IV. FEEDBACK OF USERS**

27. A big advantage of the Internet is the close relation between suppliers and users. The feedback we get from the users of our programme is in the form of suggestions, criticism or often praise for the presentation of our data. A specific survey among users is aimed at providing information about, for instance, the users' technological basis and the data they are mainly interested in. Since March 1995, access to the Internet have been additionally

analysed for time intervals and computer addresses. Analysing access and electronic mail over a longer period yields very important information about the strong and weak points of the stock of data offered, thus helping to provide a user-oriented service in the Internet that will be subject to continuous and dynamic enhancement.

#### **IV.1. Statistics of use**

28. The statistics of use have substantiated the necessity of using the internet for distributing statistical data and facts. The result indicated that the information is mainly retrieved during the usual working hours and leisuretime use rather plays a minor role at least as far as technical Internet information is concerned. This statement has been confirmed by an analysis of requests in a breakdown by time (day/night) and weekdays. The conclusion is that the main users of the programme are professionals. However, the above information is subject to the restriction that not all accesses are recorded. The number of completed requests using so-called proxy servers is unknown.

29. Based on national codes in the computer addresses the national origin can be detected for about 90% of accesses. This is important as regards language priorities for the information stored. So far, the emphasis is on German, with about 6 in 10 request's coming from German-speaking territories. About every tenth request originates from the USA. The remaining accesses to the information stored in the World Wide Web are really world-wide as they come from every corner in the world: e.g. Australia and New Zealand, China, Malaysia, Uganda and Brazil.

30. The detailed statistics of completed requests may also serve to analyse the individual demand for the different subject-related parts of the programme in order to make respective adjustments in the future. Amounting to one third of the total demand, the emphasis of information requests is clearly on basic data, followed by the press service (just above 20%) and general information (e.g. About us, Helpline, Bookstore, and Mediastore; just under 20%). The remaining requests concern the time series service and indicators whose shares are still rather modest. However, this situation may change as the latter programme parts were added only recently.

#### **IV.2. User survey**

31. A user survey developed at the Federal Statistical Office has provided additional information according to which, for instance, the interest of users focuses even more clearly on basic data. Over a period of four months 660 persons participated in the survey. A total of 63% of them indicated that basic data were of special interest to them, while 19% of them showed particular interest in general information and 9% each were especially interested in press releases and other parts of the programme. Though the survey cannot claim to be representative, its correspondence with the results of the user statistics suggests that specific attention should be paid to the quality of the stock of basic data offered. The quality and structure of the

information we offer in the Internet was qualified as "good" to "very good" by 72% of the persons interviewed.

32. In addition, the user survey provides detailed information about the purposes of using the statistical data offered in the Internet. Apparently, professional purposes come first (40%), followed by scientific purposes (24%), private/other purposes (19%) and educational purposes (17%).

33. It will be of interest to observe the development of the share of "loyal users" of the information we offer in the Internet. So far, every fifth person participating in the survey indicated that he/she had made use of the programme for several months, and every third person considered herself/himself a regular user of the programme.

#### **V. INTERNET IN THE CONTEXT OF THE FEDERAL STATISTICAL OFFICE'S MARKETING STRATEGIES**

34. Applying any new medium in the publication sector - whether diskettes, CD Roms, on-line services or the like - makes it necessary to think about the principles of the publication and distribution system and the marketing strategies. Questions arise such as: Will the new medium compete with the conventional ones? Will it replace them? Will it attract new groups of users? How will traditional users respond to the new medium? How to adapt prices to the changed conditions?

35. One of the aims of offering statistical data in the Internet is to attract new groups of users. It would still be too early to make a final judgment as far as this issue is concerned. However, user statistics and responses received by electronic mail have indicated that many Internet users did not receive statistical products via other media in the past. The number of completed requests increasing from month to month apparently proves that the medium has been generally accepted to an extent unprecedented over such a short period of time.

36. By now, the stock of data offered in the Internet could become a serious competitor to conventional publications in the near future since our office will continue to go the new way of electronic commerce. There are much uncertainty in this area of doing business but the statistical offices have the ideal products in digital forms to generate money in the middle or long run. Though the number of users of the Internet are rapidly growing, the conventional forms of publishing data will have to be maintained.

37. A large number of e-mail inquiries have referred to the question whether the Statistical Yearbook could be published in the Internet. Such an extension of the stock of statistical data offered would exceed the limits of a free provision of services. The information published in the Internet is meant to attract the attention in particular of new users so that they will also get interested in publications that are not provided free of charge. The free access to data is just to ensure that the general public will be provided with

basic statistical information. In line with that principle, the retrieval of basic data, slightly detailed economic indicators, press services and service pages from the Internet is free of charge, while the retrieval of information from the time series data base is not. Determining the price of the latter is based on the procedure used so far in establishing prices for the transmission of time series from the statistical data base. The subscription price amounts to not more than DM 50 which is within the means of students as well. However, considerations to provide access at a reduced price for research facilities have not yet been finalised.

## **VI. INTERNET AS AN INSTRUMENT FOR MODERNISATION**

38. Since the data stored in the Internet can be offered worldwide to any number of persons at the same time (except when lines get blocked due to excessive access), the information services of the Federal Statistical Office are expected to be relieved noticeably. The extent of answering inquiries usually on the phone or in writing, which requires a lot of staff and is restricted to the normal working hours, can consequently be reduced.

39. Recording the search queries in the Internet has shown that certain subjects are of specific interest. Every tenth keyword query refers to the subject of prices (incl. cost of living and inflation). As experience has shown, that kind of queries is particularly often made by lawyers dealing with stable-value clauses. For this reason, a specific stock of data regarding that subject has been offered in the Internet. Hopefully, inquiries, the replies to which require comprehensive work of the information service staff, will thus be covered by the Internet programme in the long run.

40. However, profiting fully from the relief potential will be possible only upon the establishment of an Intranet. Functioning as a 'local Internet' for internal communication purposes, the Intranet can for instance handle any e-mail matters such as receipt, dispatch, allocation, delivery without any time delay and manual interference. In addition; workgroup computing and document management tools can be advanced and perfected by means of that basic technology.

41. In another respect, too, services requiring extensive staff capacities can be reduced. In view of the fact that not the whole range of information that can be potentially extracted from statistical surveys can be stored and kept ready in processed form for any needs that may arise, a data base is required in which basic data are stored together with metadata. Many users are interested in the availability of electronic data media in order to process the information obtained on their own PCs with the help of standard or special software. The Internet, offering a user-friendly access to data via meta-information, reduces the necessity of personal advice regarding the selection of data and the use of printed information. The HTML, better XML, standards applied in the World Wide Web serve as a basis for providing a large group of users with easy access to the meta information supplied by statistical offices. XML offers the opportunity to establish an intuitively operable user interface which is very well suited for information presentation.

Consequently, a potential user needs neither to install special software on his computer nor to operate a special data base interface.

42. XML standards will probably also help to simplify the publication procedure which is rather complicated at present. The main point in all related considerations is the availability of an electronic data base with interfaces to the Internet and Intranet. That data base could form the basis for producing various publications and providing information.

43. The Federal Statistical Office and the statistical offices of the Lander have been cooperating in the development of a common new Statistical Information System (GENESIS) aimed at fulfilling the above requirements. Now the first GENESIS version will be completed. It will be used for providing information and compiling publications, thus forming the basis of an information system that will be accessible on-line via the Internet for professional and also occasional use. It is based on the multidimensional data model and it gives through it consequently implemented meta-data access a solid basis for the future. A precondition still to be fulfilled is the development of an additional user interface which will be capable of using the opportunities offered by the World Wide Web service.

44. The Internet will be used to an ever greater extent, it also serve to gather data in a more efficient way by utilising electronic questionnaires. The solution of confidentiality problems and the development of suitable interfaces to internal computing systems has made strong progress. In the Federal Statistical Office we have a pilot in the internal trade statistic, where we offer a JAVA applet to report to the statistical office. The project is named W3STAT and it uses all the latest technologies in the Internet field. International standardisation organisations especially the object management group are presently examining ways of structuring statistical data with XML syntax and using CORBA technologies for the ways of electronic data interchange. Also the establishment of a virtual private network between statistical offices could ease the workflow.

## **VII. SUMMARY AND PERSPECTIVES**

45. The globalisation of markets and ever more intensive international competition have caused an increasing demand for subject-related information on the one hand and brought about new requirements for the dissemination of information on the other. Providers of information are faced with rapidly developing communication and information technologies. In this context, the idea of a world where anybody can be present in any place at any time (mentally rather than physically) is called to one's mind. Visions of that kind, which some years ago were only considered realistic by some few people believing in rapid progress, have become generally accepted today.

46. However, it is difficult to envisage concrete tendencies in development. Further trends in. Internet application depend on a wide range of factors such as the development of even more user-friendly and self-explaining software,

the capacity of transmission networks, the development of access speed and naturally also prices and, last but not least, such factors as fear of social isolation and the like. The medium as such is enhanced at an almost incredible pace, the keeping up with which requires an immense effort.

47. With regard to statistical work at an international level, it can be assumed that statistical organisations will come-closer as well. The requirements relating to comparable results will become stricter. The methods of data collection will need further harmonisation to ensure that the results obtained will correspond with each other. The opportunity, just by a mouse click, to jump from the information offered by one provider to the information provided by another one will require identical definitions for identical terms. The regulations of the European Union on short-term economic indicators and structural business statistics have achieved great progress with regard to harmonised definitions and methods of data collection. That kind of work should be intensified at the international level and extended to other statistical areas.

48. A new type of websites, the portals, are offering information, services and products in a very comfortable and complete way. We have to figure out how such a central point for statistics can deliver such core features like

- web searching for statistical information;
- news about statistics;
- reference tools;
- access to online services;
- some communication capabilities;
- downloadable software for statistical questions (mobile code).

49. The start of such a project should not be local. It must be global to meet the objectives of all users of statistical information.