UNITED
NATIONS

E

Economic and Social
Council

Distr.
GENERAL

CES/AC.71/1999/14
29 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 USE OF INTERNET AT STATISTICS SWEDEN TO COLLECT AND DISSEMINATE STATISTICAL DATA

Submitted by Statistics Sweden[1]

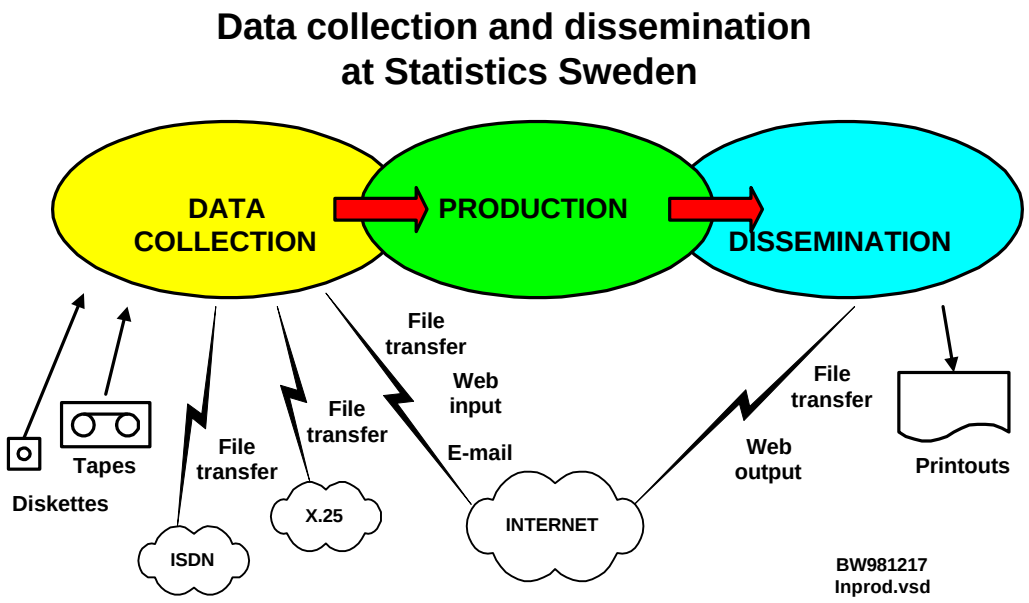
I. INTRODUCTION

1. The impact of Internet was recognized in 1994 at Statistics Sweden (SCB). Since then, a successive effort has been made to use the medium increasingly for data collection as well as for statistical dissemination. The author - Björn Walters - has been project leader for a couple of the ”Internet projects” and for the ”Data collection project” and is responsible for data communications at SCB.

II. THE DATA FLOW

2. It is natural to describe a computerized process in terms of input - processing - output. In our case, we will call these steps data collection - production of statistics - dissemination. At SCB, these steps are materialized in three separate data networks, which are connected to each other by certain security solutions. See figure below.

[1] Prepared by Björn Walters.



II.1 The data collection network

3. The data collection network gathers all electronically transmitted data which comes from different data sources and passes the data further into the production network. Examples of transmission types are e-mail with enclosed files or file transfers using the ftp-protocol or the https-protocol. In fact all three types of transmission are built upon Internet techniques. Of course there are still other ways to send in the data - on diskettes or on paper. However, using these media will soon be history. The networks used today for data collection are the Internet, the telephone network, ISDN and X.25. The last two are still used for historical reasons, but will be replaced by the Internet within a short period of time. The telephone network is used by some communities using FirstClass while others use FirstClass in combination with the Internet. Internet will probably be used 100% here as well in the future.

4. Within the data collection network there are different servers, which either collect data themselves from data sources outside SCB or to which data are transmitted from the data sources. From a security standpoint the fetch principle is preferable. Security is assured using firewalls which filter non-acceptable protocols and allow traffic to and from certain servers on the inside only.

II.2 The production network

5. In the production network there are ”target servers”, which are the final destination of the raw data. From here on, the data which mainly arrives in the shape of tagged flat files is processed and put in SQL databases for further processing and statistical treatment. The statistics are then forwarded to the dissemination network, either in the form of flat files for specific customers or as input to the official statistical databases on the Internet.

II.3 The dissemination network

6. The dissemination network is completely Internet related. It was established in October 1985, when we first introduced our website. On 1 January 1997 we presented the official statistical databases on the Internet and during the autumn of 1998 we introduced a web hotel as well as a database hotel for statistics. The official databases on the Internet are updated frequently according to specific procedures. The databases consist of SQL databases and of ready-made PCAXIS-files which can be downloaded to the PC. One can also extract data from the SQL databases in the shape of ASCII files, which can also be downloaded to a PC and to a Mac or a Unix work station.

7. I mentioned earlier that flat files could be forwarded to the dissemination network. These files are placed in specific catalogues on the web hotel, where the customers can fetch their files.

III. EXPERIENCES

8. Some experiences that could be of common interest are mentioned below.

III.1 Data collection

9. Of course, one tries to standardize and to rationalize to reduce the number of input possibilities, but this is not so easy. One has to accept certain variations due to constraints and demands of data providers, i.e. delivery demands.

10. The data collection network must be very carefully documented so that the people responsible for managing the data collection can react accordingly when, for instance, a server fails. There should be backup machines and the staff should know what to do in various situations, etc.. The documentation work would seem straightforward, but we have had some bad experiences due to having outsourced the network management a long time ago and to having changed the outsourcing company, etc..

III.2 Production

11. The production staff also have demands. They might want the incoming data in a certain way, untagged for instance. They want a notice when data has arrived. They should be informed when something has gone wrong, etc.. Such notification will be made by e-mail. Close cooperation between the ”input staff” and the ”production staff” is very important. Awareness of the production processes is essential when designing the input network.

III.3 Dissemination

12. More and more data are disseminated through Internet. It is recommended to use correct publishing procedures for the dissemination, i.e. the more automated the dissemination process, the better. SCB decentralized the publishing of information on the website from the very beginning and that has worked out well. The method was to use internal servers to build up information and to propagate it to the external website when checked and approved. When it comes to the databases, a more thorough control is necessary. But also here decentralized responsiblility is the only solution. By means of internal database servers the different statistical programmes check that their data looks as expected before it is forwarded to the ”dissemination servers”.

IV. THE FUTURE

13. Several projects are underway at SCB.

- One to build the Swedish Statistical Network, combining statistics from the different stastistical authorities (SAM) in the country. This information will be linked by using hyperlinks on webpages. The information will be free of charge on the Internet.
- Another project will lead to an extranet within which the SAM can cooperate and exchange exlusive information.
- A third project aimes at putting all the SCB publications free of charge on the web. Publications will still be obtainable on paper, but the printing costs will be charged.
- A fourth project is the first at SCB to use web forms for interactive input of statistics from the Internet. This technique is expected to be widely used in the future.

14. In Sweden there is a project underway which aims at introducing a system - SHS - by which it will be possible to retrieve and disseminate information between authorities and other organizations in a standardized and secure

manner on the Internet. SHS stands for Dissemination and Retrieval System. The project is in the aquisition phase.

15. The impact of the market is an essential factor. More and more organizations abandon special networks in favor of the Internet. Continued adaptation to market requirements, i.e. Internet, and related techniques will therefore be a wise strategy. Adaption to standard software development tools and standards as much as possible will be the strategy for SCB. Adaption to country specific standards, e.g. SHS mentioned above, is necessary.

16. The picture of the data flow shown in the beginning of this paper will change in future to look like the following scheme.