

SEMINAIRE

E+; 3=! C

SEMINAR

STATISTICAL COMMISSION AND
 ECONOMIC COMMISSION FOR EUROPE



Distr.
 GENERAL

CONFERENCE OF EUROPEAN
 STATISTICIANS

CES/SEM.43/30
 7 April 2000

ENGLISH ONLY

Seminar on integrated statistical information
 systems and related matters (ISIS 2000)

(Riga, Latvia, 29-31 May 2000)

Topic IV: Improving data dissemination strategies

**IMPROVING DATA DISSEMINATION STRATEGIES IN THE
 CENTRAL STATISTICAL OFFICE OF POLAND**

Contributed paper

Submitted by the Central Statistical Office of Poland¹

I. INTRODUCTION

1. As of March 1997 two units are responsible for statistical data dissemination in the CSO: the Data Dissemination Department and the Information and Publishing Office. The Data Dissemination Department is responsible for providing data for: management, administrative duties, and international cooperation. The Information and Publishing Office passes the information on to users, including the media (data dissemination) and prepares market analysis - analysis of the need for the given statistical data.

2. The task of bibliographical data dissemination is given to the Central Statistical Library (CSL).

3. The most commonly used means of statistical data production and dissemination in the Central Statistical Office (CSO) are:

- interview, including by telephone,
- forms, including ones distributed by fax,
- e-mail,
- Bulletin Board System (BBS) available via modem,
- off line: CD ROM, disks,
- Intranet,
- Internet.

4. The basic requirements for the disseminated information are: accuracy, appropriate data structure, data addressed to the desired recipient, and available to all recipients at the same time.

1 Prepared by Halina Agnieszka Stegawska.

5. According to the Polish statistical law (Act of 29 June 1995) all users have the right to equal, equivalent, and simultaneous access to statistical data. Internet is one of the major forms of data dissemination used by the CSO's information system. All data processing resources being inside CSO's Intranet are closed by the "firewall system" to external users - only servers with public databases, being connected in front of the "firewall system", are accessible via Internet or via switched telephone lines (BBS).

II. CURRENT STATISTICAL DATA DISSEMINATION SITUATION

6. One of the Data Dissemination Department's responsibilities is to prepare Information packets. These packets include:

- yearly and quarterly statistical data bulletins prepared together with other departments,
- choosing the information to be made available on the www pages,
- creating the actual www pages,
- including the www pages of Voivodship Statistical Offices (VSO), on their dedicated Internet servers,
- coordinating the process of information presentation by the CSO.

7. According to the guidelines of the Economic Commission for Europe, the disseminated data should:

- be precise and given without bias,
- contain no errors,
- be current,
- be in an easily comparable form and therefore save the user's time,
- be presented in a clear, and easily readable way,
- be made available according to a release schedule, notification to the users of any delays,
- include an appropriate degree of data analysis,
- be accompanied by definitions of terms used, and classifications,
- be available in the national language as well as English.

8. Currently the CSO and all the VSOs have their own Internet pages, made available by dedicated servers. The Internet pages of the CSO contain an Information part and a Data part. The Information part contains:

- statistical laws and regulations,
- the organization structure of the CSO, the schedule of statistical research for the current year,
- guidelines for making the data available,
- addresses of the VSOs,
- up-coming publications of the CSO,
- monthly information about current CSO publications,
- information about REGON and TERYT registers,
- information about the CSL,
- promotional materials.

The Data part contains:

- announcements by the head of CSO,
- current information - preliminary results,
- monitoring,
- current statistical bulletin,
- maps with statistical data,
- data according to the IMF Special Data Dissemination Standard - includes basic economic indicators, tax/budget, financial, and external trade data,
- Local Data Bank which contains 1200 characteristics of cities, voivodships, and Poland as a whole.

The address of the Home page is: <http://www.stat.gov.pl>.

9. The VSO Internet pages also have two parts - an Information part containing:

- The organization structure of VSOs,
- addresses and telephone numbers of organization units within VSOs,
- administrative sub-divisions of the voivodship,
- information about what is available at the local Information centre,
- guidelines on how to register a business in the REGON register,
- schedule of upcoming publications, and summaries of published material,
- promotional material;

and a Data part with the basic indicators and data for the voivodship.

10. The Information and Publishing Office has the following responsibilities:

- providing domestic and international users with information support, including consulting, clarifications, and advice;
- accepting orders for:
 - the CSO's statistical data which is not normally published,
 - numerical data - production and revenue according to SWW (Systematic Product Register), and PCN (Polish Nomenclature for International Trade),
 - Data from the the National Official Register of the Nationalized Industries Units KRUPGN (previously REGON),
 - basic information about PKWiU (Polish Classification of Products and Services);
- automatic statistical information:
 - price index and average incomes;
- Press Office:
 - information support for mass media,
 - organization of press conferences.

III. IMPROVING DATA DISSEMINATION STRATEGIES IN THE POLISH CENTRAL STATISTICAL OFFICE

11. Due to the need for improvement in the system of data dissemination and presentation, a concept of Statistical Information System (SIS) was developed based on a 'data warehouse' technology. A common database is currently under development.

12. The SIS is one of the information systems of the public administration and it provides:

- A spatially and temporally organized database of information: data collection techniques, sources, storing places, and delivery points;
- Set of rules for information collection, processing, saving and storing. It will also outline how to deliver this information in a given time and form to users;
- Set of data elements which is coordinated and of the desired structure;
- Ease of use, despite the inherent complexity of having a large number of interacting system elements.

13. Among its other uses, the system will provide the information for databases of the Statistical Publishing Establishment; Catalogues of tables, maps, charts, frequently ordered items; and information for many international organizations. The amount and scope of information presented on the Internet and through the Bulletin Board System (BBS) will be increased.

14. Making the data available in a 'warehouse' type common database enables the user to:

- Search and browse through data using any 'key words',
- Print or save the data to a disk, CD-ROM, or another file,

- Perform the standard statistical analysis - then print and save,
- Create a table based on individual needs (the columns and rows will be user defined) - then print and save,
- Make graphical presentations by creating maps, charts, and graphs - then print and save,
- Aggregate data based on a given algorithm for processing,
- Easily use the archive data.

15. The key aspects of the system being developed are:

- Flexibility in grouping the data,
- Many different information sources,
- Flexibility in the result presentation, including the ability to create charts, maps, etc.,
- Many easy ways to create graphs through setting different parameters, such as the range of presented data, type of graph, layout, title, axis labeling, and ability to include remarks.

16. Creating a system with diversified and vastly spread inputs and outputs is a task which must be realized over an extended period of time. It is necessary to divide it into many phases. The first such phase was limited to integrating to SIS those elaborations which deal with the economy on the national level. Agricultural information pertaining to individual farms was excluded because there is a separate agricultural database being built.

IV. CENTRAL STATISTICAL LIBRARY

IV.1 Current situation

17. The contents of the Central Statistical Library are available through the Internet using an Automated Library Expandable Program (ALEPH). It provides Internet access to a retrospective database (6000 records) as well as supplementary current bibliographic descriptions, CSO publication database, statistical yearbooks and journals.

18. The system is of the client-server type, based on an ORACLE database. It is characterized by a user friendly, graphical interface and hypertext capability.

19. The system consists of modules which handle its different tasks, such as:

- Cataloguing, with user adjustable format,
- On-line Public Access Catalog (OPAC) searching, done by index browsing or by typing in the author's name, title, subject, ISBN number, or key words used in the summary,
- Controlling the vocabulary and self-acting composition of the Thesaurus,
- Controlling the amount of information available,
- Receiving and dealing with mail for orders, returns, etc.,
- Making information available in real-time conversation mode,
- Checking the publishing of periodicals.

20. Integrating the Central Statistical Library to the Internet gave the inside users of the CSO the ability not only to browse through this library's database, but also to browse through all library databases available on the Internet. The OPAC catalogue is available through any computer with an Internet access through the World Wide Web. To help Internet users move through the OPAC catalogue, special information screens were created and implemented in the ALEPH. As more functions are added, the user will be able to do more on line; for example, order publications or copies. More information about ALEPH and the database of the Central Statistical Office is available on <http://statlibr.stat.gov.pl>.

IV.2 Development strategies

21. Using the experience gathered while working on the project, and the experience of other Polish and European Information Centers such as ones from Italy, Germany, Denmark, Czech Republic, or Hungary, a development plan was designed for making the library information available.

22. Work to increase the number of Internet functions for data delivery is in progress; between August and December 1999, 7630 Internet sessions were registered in the CSL database. Including international requests, approximately 100 detailed e-mail cross-reports were made for example, about forest fires in Poland over the last century. The goal of the work is to make more options available when ordering and opening personal accounts, and also to decrease the time necessary to make a detailed report.

23. Cooperation of CSL with other European statistical libraries is planned using the Internet. In 2001 CSL is organizing a Seminar on Experience and Potentials in Development of the International Cooperation of Statistical Libraries.

24. A new version of the system - the ALEPH 500 - is being integrated. This version was built based on Oracle's procedures for data base management. As compared to the currently used ALEPH 300, the new version expands the range of the internal library as well as end-user procedures.

V. SUMMARY

25. The core principle of statistical data dissemination is that all users have the right to equal, equivalent, simultaneous access to statistical data. The basic principle of developing data dissemination process in the CSO of Poland is to make the best use of all media available. Internet is the fastest and most convenient media for data dissemination. It permits the transmission of the right product to the right people in the right way, and at the right time.

26. The SIS project, based on an integrated data 'warehouse' system, was started in order to further develop the methods and procedures of data dissemination. Creating such a diversified system with a wide range of inputs and outputs, is a task, which can only be realized in phases over a period of time. The first completed phase was limited to integrating only those applications, which deal with the national economy. As phases of SIS are completed, new technical solutions must be also made, such as equipment purchase and other projects.

27. The experiences of other statistical institutions, which have already been through this process, will be very helpful to us, both from the technical and methodological perspectives. Those experiences will allow us to shorten the time needed to integrate our statistical information system with countries of the European Union.