Distr.
GENERAL

CES/AC.71/2001/14
20 December 2000

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

COMMISSION OF THE EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE/Eurostat Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 14-16 February 2001)

Topic (ii): Metadata modelling and terminology issues

USE OF NETWORK TECHNOLOGY IN THE CZECH STATISTICAL OFFICE: CURRENT STATUS AND FUTURE PLANS

Submitted by Czech Statistical Office [1]

CONTRIBUTED PAPER

I. DEVELOPMENT OF TECHNOLOGICAL SUPPORT FOR STATISTICAL PROCESSING IN THE CSO

1. In the Czech Republic (CR), like in other countries, the technological environment in which statistical surveys used to be processed for a number of years was characterized by the prevailing use of mainframes, mainly due to the volume of the files handled and the outputs produced. All calculations were carried out centrally on the mainframes in the Prague head office and the results were presented on paper only, in the form of printed outputs. Data preparation was also centralised by transferring from statistical forms onto media that could be computer processed. The media were transported physically, as no transmission lines were used for data transfer.

2. In 1991 the Czech Statistical Office (CSO) commenced its gradual transition to a network structure. The last mainframe (Cyber 180) was dismantled in 1996 and fully replaced with a network of

[1] Prepared by Jan Matejcek.

workstations, more widely distributed and interconnected. The CSO started to use permanent data lines whose data transfer rates increased gradually.

3. The level of hardware and software achieved in the CSO at the beginning of the year 2001 is characterized by the following: the basic processing capacity in the CSO's computer network is provided by Compaq servers (ranging from ES40 to GS320Wildire in type) which work for the most part under a cluster mode. The servers fulfil communication, application, data, back-up, distribution, presentation, installation, security, administrative and training functions. They are united on the platform of the operating system Unix and the database system Oracle. Data are stored by means of external RAID arrays (ESA and EMA12000) 4TB in capacity, backed-up by 22TB tape libraries ESL and archived using 2TB magneto-optical libraries Plasmon.

4. Personal computers with 32-bit-based operating systems Microsoft Windows NT and Microsoft Windows 2000 Professional are used for client's stations in the CSO. There are over 1 800 of them in the office and virtually every employee works on his/her computer incorporated into a network (Ethernet). The CSO practically stopped using Corel SW (WordPerfect, Quattro) and began to work with the Microsoft Office SW (World, Excel).

5. The LAN networks in Prague and regional towns make it possible to operate at a data transfer rate of 100 Mbit/s; WAN lines are interconnected in an ATM/FR network 128 and 256 kbit/s in transfer rates and with encryption. Connection to Internet occurs at a rate of 2 Mbit/s as a rule and can be accelerated almost arbitrarily, if need be. The public segment of the network, which provides all communication outside the CSO, is rigorously isolated by means of Firewall technology. All of the data transfer lines are currently fixed lines. Mobile telephone networks are available at present only to some employees, but they are expected to be widely used in the near future.

II. CURRENT METHODS OF PROCESSING IN THE CSO

II.1 Statistical surveys

6. The processing of groups of questionnaires (industry, constructions, investments, trade and services, agriculture, forestry and the environment, prices, banking and non-production sphere, social statistics, demography and wages) is split and assigned to seven nation-wide data processing departments located in regional towns. Exhaustive or sample survey questionnaires either on paper or partly in electronic form are collected in these departments from respondents from all over the country to receive initial processing, checks and corrections in the local LAN networks. The data are then transmitted via WAN network to Prague to be centrally treated in a database system. The outputs on paper, Intranet and Internet are available to local and remote users.

II.2 Field surveys

7. In statistical surveys calling for respondents to be assisted by statisticians (surveys on consumer prices and household budgets, labour force sample surveys, censuses, etc.), CSO employees help respondents fill in questionnaires. Partial recording into notebooks with the Blaise system and optical character recognition are now used for the labour force sample survey and the agricultural census respectively. The questionnaires are loaded into PCs at detached units of the CSO's regional representations, established in district towns, and sent off to Prague (consumer prices to the town of

Hradec Králové) for central processing. Again, outputs from central processing are available in publications and on Internet/Intranet.

II.3 Population census

8. The housing and population census to be carried out in the CR in 2001, normally every ten years, is a challenge for the CSO in this period. As the population register of the CR is not of the quality desired, the technology used this time has to rely on optical recognition of characters in over 15 million forms. The technology will use a network environment at both the input (a network of workplaces established to make checks and rectify errors in the forms and the optical character recognition) and the output (where dissemination of the results by means of Internet and CD-ROMs is a priority).

II.4 Elections

9. Amendments to relevant laws stipulate that the CSO assumes full responsibilities for the processing of electoral results since 2000. Partial results are brought on floppy disks (and on paper to a small extent) from 14 000 constituencies to 500 authorised local authorities. The authorities and the central processing workplace at the CSO in Prague make up a temporary FR/ATM network 64 kbit/s in data transfer rate. The data are transmitted along this network from local authorities to Prague for continuous processing. The whole course of the real-time processing is available on Internet, which is appreciated by mass media and political parties. The processing is very fast and easy to check due to its openness. For instance, the second round of the elections to the Senate was processed within 3 hours after the polling stations were closed.

II.5 Administrative activities of the CSO

10. Given the fact that virtually all employees of the CSO have their computers connected to a common network, communication between them is to a large extent electronic. Communication in writing is becoming exceptional and was superseded by e-mails in the GroupWise system. The still prevailing shared files on reserved disks are transferred into an Intranet environment. Currently, the Intranet is composed of the following divisions: news, directory, tasks and applications, documents, management, publications, services, and discussions.

III. CHANGES RELATED TO NETWORK TECHNOLOGY UNDER PREPARATION

III.1 Growing volume of questionnaires completed electronically

11. The current share of electronically completed questionnaires, which ranges from 1% to 10% now, depending on the type of questionnaire, does not correspond to the number of computers available to respondents in the CR. Virtually all businesses subject to reporting duty have PCs and 90% of them are connected to Internet. There are several causes behind this disproportion, one of them being the low stability of statistical surveys during the period of harmonizing Czech statistics with the EU. When the harmonization has been finalised, the respondents will be capable of efficiently connecting the production of required statistical data to their information systems.

12. The expected cut down on Internet-connection costs will make it possible to move from today's mode of pre-defined completion of questionnaires and their subsequent e-mailing to an interactive mode. Currently, the respondent uses WWW technology to choose the required questionnaire (file) and opens it on his/her PC, where a respective variant of the interactive programme is generated, including checks,

nomenclatures and explanatory notes. After the filling-in stage is completed, the interactive programme produces a data file to be supplied to the CSO by e-mail or on floppy disk. This technology is already in use for all statistical questionnaires. The interactive completion of questionnaires on a WWW-site and the application of electronic signature after the certification authorities have done their job and authentication servers are active will mean that the collection of signed paper questionnaires will be almost obsolete in years to come.

III.2 Abolition of the CSO's district units

13. Today's organizational structure of the detached units of the CSO's regional representations follows the traditional structure of districts which is to be abolished in the Czech Republic on 1 January 2003 and replaced with authorized municipalities. However, the technology of field surveys will have to be modified so that this statistical district-based structure can be physically abolished (offices, local PC networks, data transfer lines).

14. Statistical surveys of consumer prices are being adjusted to the technique of handhelds, labour force sample surveys will be fully equipped with notebooks, optical character recognition or mobile computer technology is under consideration for household budget processing, and censuses will be processed by means of either optical character recognition or keeping constant registers. All transmissions to central databases are supposed to be effected via mobile telephone networks (GMS, GPRS, UMTC). Pilot projects put in place for these statistical surveys in the field will also include full harmonization with the EU and new requirements for regional data structures.

III.3 Growing share of outputs on Internet

15. Almost all outputs produced in the CSO are available today not only on paper, but also on Internet and CD-ROMs. However, outputs of temporary validity or those designed for a rather narrow spectrum of users, which are offered on paper, will be gradually abolished and disseminated using electronic media only. This approach has proved very successful in supplementary activities such as the processing of all kinds of elections. The huge amount of housing and population census outputs will use these media from preference too.

16. The current structure of the CSO's fundamental presentation (excluding elections and housing and population census, which have their own WWW-sites) consists of the following parts: news releases, hot data, topicalities, services of the CSO, classifications and nomenclatures, methodology, business register, submission of questionnaires, electronic data collection, contacts, vacancies, publications, and analyses, broken down by individual cross-sectional and industrial statistics.

17. In 2001 the structure will be adjusted to the common government backbone gate now being introduced, which standardizes the form of communication of the public and the central and local governments throughout the CR.

III.4 Establishment of the state administration communication infrastructure

18. Unlike the internal infrastructure at individual ministries, no backbone communication network for the Czech state administration as a whole (the so-called Extranet) has been established to date. The intention of the state information policy is to build up such a network fast. The network will rely on defined shared

information interfaces, an integrated state administration gate and commonly used public information terminals permitting citizens to access the network. This will make it possible to mutually use administrative sources within public administration, while respecting the principle of entering each item of information only once.

19. Of crucial importance for the work of a statistical office are basic nation-wide registers such as the population register, business register, real estate register and territorial identification register. The CSO will significantly participate in administering and operating these registers. The Office is to be responsible for a solution to and the administration and operation of the basic business register, will participate in solving the basic territorial identification register and operating it, and will participate in elucidating the basic population register.

III.5 Regionalization of statistical information

20. The requirement to provide information for the management of the development of regions at all NUTS levels calls for the need to dispose of regional cross-sections for most statistical surveys. In a network technological structure this task is not too difficult to cope with from the technical viewpoint. However, it is more difficult to solve where the subject-matter and funding are concerned. Since demographic and social statistical services have been working with detailed regional structures for a long period of time already, all relevant data related to NUTS2 and NUTS3 levels will be available after the completion of local and kind-of-activity units of businesses. However, individual data protection is increasingly more difficult to ensure for NUTS4 and NUTS5 levels, as these territorial units are comparatively small in the Czech Republic.

III.6 Connection to international institutions

21. Individual connections of the CSO to international institutions such as Eurostat, ECE, OECD, ILO, UN and the like are not envisaged. Such a solution would be demanding in terms of funds, though it could be implemented virtually at once. A solution based on the interconnection of the Czech Republic’s common state administration network with these supranational institutions would seem to be more effective. Of course, until such a system is implemented, the use of the public network Internet will be expanded.