

IT management and database development strategies and current information on the IT system

I. CSO ICT organization and management

CSO ICT (Information Communication Technology) system creates a foundation for an execution of essential tasks of Polish official statistics, i.e. carrying out of statistical surveys and disseminating their results, as well as maintaining two national official registers: TERYT – concerning administrative/territorial division of the country and REGON – containing entities of national economy.

ICT organization and management is based on activities and input of the following units:

- 1) Central Statistical Computing Centre, being the main player, executing more than 70% of IT tasks of Polish official statistics,
- 2) CSO Divisions,
- 3) Some centres of competencies at regional Statistical Offices, which operate at the voivodship level,
- 4) ICT Units, Interview Survey Units and Register Units at the 16 regional Statistical Offices (SO),
- 5) IT groups at the Central Statistical Library and Statistical Publishing Establishment.

Essential task of the **Central Statistical Computing Centre (CSCC)** in Warsaw (having also two Branches: at Radom and at Łódź, altogether 300 employees) is to provide Polish official statistics with ICT services for statistical data collection, processing, compiling and dissemination, as well as maintenance and development planning of ICT infrastructure. These include the following main activities:

- Planning, development, programming, implementation, running and maintenance of approximately 200 data processing systems servicing Polish official statistics according to the annual statistical survey programme of the official statistics (in these: maintaining of the REGON and TERYT registers), collaborating on these with the CSO Divisions and related units of the regional Statistical Offices.
- Storing, running and giving information services from the developed, by CSCC together with the CSO Divisions, national statistical databases including metadata base and data files. Planning and implementing of the statistical data protection by means of ICT solutions.
- Running and administering the Polish official statistic corporate Wide Area Network with Internet access and CSO Web sites and e-mail services, as well as technical maintenance and administering of the CSO Headquarters LAN and all ICT equipment on the spot.
- Researching on development trends on ICT statistical systems, concerning infrastructure, technology, operation and implementation ways, as well as planning and carrying out of ICT investments for the Polish official statistics.

CSO divisions directly involved in the production of statistics:

- Co-ordination of Statistical Surveys Division,
- Analyses and Regional Statistics Division,
- Dissemination Division,
- National Accounts and Finance Division,
- Business Statistics and Registers Division,
- Social Statistics Division,
- Economic Statistics Division,
- Services Statistics Division,
- Agriculture and Environment Statistics Division,

are main partners of the CSCC. They prepare project guidelines for planning, development, programming, implementation, running and maintenance of data processing systems and databases to be servicing Polish official statistics. CSO divisions have also small groups of IT staff supporting execution of division tasks.

Centres of competencies, being established at 16 regional Statistical Offices (SO), carry out assigned to them selected country-wide tasks of the Polish official statistics. Tasks of part of these centres are involved in ICT services. Such are the following:

- Centre for Regional Accounts at the SO at Katowice,
- Centre for Development of Statistical Data Processing Systems and Databases at the SO at Katowice,
- Centre for Development of Statistical Data Processing Systems and Databases at the SO at Olsztyn,
- Centre for Administrative Data Sources at the SO in Warsaw,
- Centre for Regional Statistics at the SO at Wrocław.

SO ICT Units, Interview Survey Units and Register Units, managed by directors of the regional Statistical Offices, are responsible for the following tasks:

- Collecting of questionnaire-based and interviewer's data, processing, compiling and storing data related to their voivodship territory within the annual statistical survey programme of official statistics, using the unified application software, being developed in general by the CSCC, as well as executing similar tasks when SO conducts specific surveys for regional or local needs,
- Collecting data by interviewers (approximately 840 interviewers in total in 16 SOs) equipped with PC notebooks having data communication facilities,
- Registering economic entities and updating the REGON register in the scope of their voivodship territory, using the unified application software, developed by the CSCC.

IT groups at the Central Statistical Library and Statistical Publishing Establishment support for main services of these units:

- running computerised library system, being accessible for querying bibliographic information by users through Internet,
- preparation of printing production of statistical publications, basing on received data files containing tables and texts prepared by the CSCC and CSO divisions.

II. CSO databases and related development strategies

Two national official registers: TERYT – concerning administrative/territorial division of the country and REGON – containing entities of national economy, being the base for survey organization and production of statistics, are carried out in the database technology.

TERYT register covers:

- system of code identifiers and names of units of the state administrative/territorial division (voivodships, poviats, gminas), being supplemented with Polish NTS/NUTS statistical units (regions consisting of some voivodships, sub-regions consisting of some poviats),
- system of code identifiers and names of localities,
- system of code identifiers and names of streets,
- system of identification of addresses (real estates, buildings and dwellings),
- system of statistical districts and census enumeration areas.

REGON register has collected all entities of national economy (approx. 5 million entities), carrying out business activity (legal persons and organisation units not being legal persons and natural persons), including also agricultural farms, as well as entities acting in administrative, social, cultural and other areas. There are registered names and address data of the a/m entities and their local units, classifications and legal forms of their activities and several other data. Address data of the business entities are compliant with the contents of the TERYT register and include its system of identifiers.

BJS statistical business register, carried out also in the database technology, is based upon REGON set and has an additional set of entity data, being updated from selected administrative and statistical data sources. It is compliant with the EU regulations. BJS register role is to support defining of sets of individual business or other entities to take part in specific statistical surveys, as well for complete surveys as for sample surveys.

SRGRiL statistical register on agricultural farms and forestry holdings, which has been established in the database technology using data of individual farms collected within national Agricultural Census 2002 and is updated using data of every year statistical surveys and of administrative data sources. It is compliant with the EU regulations. SGRiL register role is similar to the BJS one.

CSO units maintain several databases:

- a) with statistical micro-data and/or aggregated data and relevant metadata (named later “**internal databases**”) – for internal needs and for provision information services by CSO according to orders of external users,
- b) with statistical aggregated data and relevant metadata and time series capabilities (named later “**public databases**”) – for provision information services for external users directly through CSO Web sites.

There are maintained following **CSO internal databases**:

- **Foreign trade database named HINEX**, which had run since 1992 and has been modified in 2004, following the Poland’s accession to the EU. It contains annual Poland’s foreign trade data as of 1992 and monthly data of current year. HINEX database is updated every month by means of data files collected by Ministry of Finance from the INTRASTAT questionnaires covering trade with EU member states and from Single Administrative Documents on trade with non-EU member states. It contains also metadata base with Polish CN nomenclature of goods, list of countries and set of Polish economic entities participating in foreign trade activities. To facilitate analyses and user access there

are prepared and stored approx. 1000 statistical tables presenting the foreign trade aggregated data as of 1992. HINEX database is the source of data on Polish foreign trade transmitted every month to Eurostat, according to GESMES (Generic Statistical Message) standard using GENEDI (Generic EDI Tool) application software. Moreover there is prepared interface software giving to external users possibilities for simple querying at the CSO information centre an aggregated data on selected CN level by countries or group of countries.

- **Central economic database** has gathered individual and aggregated statistical economic and financial data of economic entities from 10 (ten) monthly, quarterly and annual surveys, as of 1990, which comprise data of large, medium and sample of small enterprises.
- **Structural Business Statistics database** – contains annual basic statistical economic data of enterprises and their local units, being gathered within structural surveys, as of 1994.
- **Database on Enterprises with Foreign Capital Investments** – comprises annual statistical data on the foreign and national capital of enterprises, balance of assets and liabilities, incomes, costs and financial results of activities, as of 1995.
- **Privatisation database** – includes state enterprises with their basic statistical data and way and results of their ownership transformation; covers annual data since 1990.
- **SME sector database** comprises annual statistical data and economic-financial indices for small (below 50 employees) and middle (between 50 and 250 employees) enterprises - since 2000.
- **Producer prices database** – contains monthly data on prices and sales value of representative items of: industrial products (as of 1986), forestry products (as of 1990), fishing products (as of 1996), construction works, transport and logistics and communication services (as of 1990), as well as monthly price indices at the representative item level.
- **Retail prices database** created in 1990 in order to store, process, update statistical data on retail prices and price indices of consumer goods and services (CPI). It contains information on prices of 1800 representative items gathered by price collectors in 209 price survey districts. Data on average prices are presented for selected towns, regions, voivodships and for the country. CPI data are stored for the last 5 years, then they are archived.
- **Statistical Data Warehouse (SDW)** is being developed since 2001. At present Statistical Data Warehouse includes:
 - data of the National Population and Housing Census'2002,
 - data of the National Agricultural Census'2002,
 - database on agricultural holdings, being compliant with the Eurostat's EUROFARM database,
 - statistical data of selected monthly, quarterly and annual economic surveys, the first ones of 2002.

Existing SDW resources are an implemented part of the Integrated Statistical Information System concept, which the second main component is to be Statistical Metainformation System (SMS), being yet under development.

CSO public databases are the following:

- **Population database called DEMOGRAPHY**, contains annual aggregated data on number of population: by sex, by groups of age, by births, by deaths, by marriages, by divorces, by internal migrations, by foreign migrations, by country territorial division (at voivodship, powiat and gmina levels), by urban and rural areas, as from 1995, being accessible to the public through the CSO Websites.
- **Regional Data Bank (BDR)** – includes annual and some quarterly statistical data and data of administrative sources concerning the demographic, social, economic and environment status of territorial units at all levels (country, voivodship, powiat, gmina), and delimited areas for different groups of users, in that: regions and sub-regions according to Polish NTS/NUTS classification as of 1994. BDR stores and updates and allows to query aggregated data by territorial units and delimited areas, being prepared on the foundation of statistical and administrative data of REGON entities, located in the considered territorial unit or delimited area, taking into account requirement of statistical confidentiality. BDR is accessible to the public through the CSO Websites.
- **TERYT register public database** – contains extract of TERYT register database with special user interface module facilitating users with browsing and querying the TERYT data, covering 3 TERYT sub-systems: on the state territorial division, localities and street names.

Referring to the DBMS being presently applied, it can be listed:

- Oracle 10g RDBMS (for BDR, TERYT and SDW),
- MS SQL Server 2000 & 2005 RDBMS,
- FoxBASE/FoxPro.

CSO database development strategies are the following:

- to improve existing public databases BDR Regional Data Bank and DEMOGRAPHY population database - to have more user friendly interface, considering main user profiles,
- to prepare Poland's macro-economic database to be public databases with user friendly interface,
- to develop a complex public database on economic statistics together with the related metadata base, having user friendly interface, being metadata driven, considering main user profiles,
- to develop a complex public database on social statistics together with the related metadata base, having user friendly interface, being metadata driven, considering main user profiles.

III. Current information on the CSO ICT system

CSO ICT system, servicing all CSO units and external users, at the 2007 is composed of:

- CSO wide area network (WAN) being a corporate Intranet, using the TCP/IP protocol and permanent virtual connections of the public data communication network (based on the Frame Relay technology, see schema below). At the CSO WAN first level, it

- links a LAN at the CSO Headquarters and Central Statistical Computing Centre (CSCC) with LANs at 16 regional Statistical Offices (SO) and at 2 CSCC branches. At the CSO WAN second level there are 16 SO sub-WANs giving SO LAN connections to the SO Branch LANs. CSO WAN is used for the statistical data file transmission within the production of statistics and further for data processing, as well as for E-mail and Internet connections. CSO WAN has only one connection to Internet via CSO Headquarters node and "Firewall system"; only CSO Web sites and public databases, installed in servers are connected in front of the "Firewall system", to be accessible for external users through Internet;
- local area networks (LANs), using the TCP/IP protocol, installed at CSO Headquarters and CSCC and at 16 regional Statistical Offices and SO Branches, in total in 83 sites having in total approx. 6000 active outlets, to which are connected all listed below servers and almost all PC workstations;
 - more than 5.000 PC workstations (approx. 60% operating under Windows XP and Windows 2000), of which 4500 PCs are connected to the LANs,
 - 470 servers connected to the LANs at the CSO units, of which there are:
 - 100 SCO-UNIX servers (Intel based) with FoxPro software involved in the statistical data collection and data editing and some processing,
 - 330 servers running under Windows 2003 Server/ Windows 2000 Server executing main data processing and data communication tasks,
 - 20 middle class servers UNIX type and Windows 2003 Server/ Windows 2000 type maintaining and running CSO databases,
 - 20 middle class servers creating a new CSO Web portal system 2007, consisting of two parts:
 - CSO Web information portal, being run under RedDot 7 Content Management System,
 - CSO Web data collecting portal, being developed applying WebLogic software tools.
 - CSO Web sites and public databases, being accessible under URL address: www.stat.gov.pl since 1995, have been visited in 2006 by 2.181.000 unique users, having 4.350.000 sessions, visiting 58.637.000 pages and executing 370.156.000 queries. It has been replaced in II quarter 2007 by the new CSO Web information portal, which has been developed through outsourcing.
 - pilot Web based statistical data collection system has been developed and implemented in 2005. It consists of electronic questionnaires, which are accessible on the CSO Web sites to the authorized representatives of earlier registered entities, to be filled in and transmitted to the data collection server. Presently there are serviced 9 electronic questionnaires on the economic statistics with the monthly and quarterly frequencies. Electronic questionnaires are utilised by some percent of reporting entity population. The first implemented questionnaire is presently applied by approx. 25 % of entities being obliged to deliver questionnaire to the CSO. New CSO Web data collecting portal, having upgraded system architecture to ensure extended operation abilities, productivity and reliability, has been developed in 2005 – 2006 through outsourcing, using experience of the pilot system. Presently it is being prepared by the CSCC staff to be implemented in second half of 2007, replacing the running pilot system and further facilitating the extension of statistical data collection through Internet, to decrease existing paper questionnaire based data collection technology.