

SEMINAIRE

E+; 3=! C

SEMINAR

STATISTICAL COMMISSION AND
 ECONOMIC COMMISSION FOR EUROPE



Distr.
 GENERAL

CONFERENCE OF EUROPEAN
 STATISTICIANS

CES/SEM.43/18 (Summary)
 2 February 2000

Original: ENGLISH

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

(Riga, Latvia, 29-31 May 2000)

Topic II: Planning and management of statistical projects

CENSUSES 2000: MAIN STRATEGIES AND IT SUPPORTING PHASES IN SLOVENIA

Contributed paper

Submitted by the Statistical Office of the Republic of Slovenia¹

SUMMARY

1. The censuses in Slovenia have been supported by up-to-date Information Technology (IT) during the last few decades. Each time when the census turned around a decade, new possibilities of IT were introduced. Over time, the use of IT has become one of the key elements in the efficient organisation of the whole census campaign. The role of IT has increased to cover almost all phases of work from Census preparation to data dissemination. The main challenges nowadays lie in:

- the integration of data from different administrative and statistical sources;
- reducing the burden of respondents;
- the integration of operational processes;
- dissemination of the results to a larger public than ever before.

2. The Statistical Office of Slovenia will organise in the near future two major censuses: Census of Population, Households and Housing (planned for 2002) and the 2000 Census of Agricultural Holdings. Both these censuses will, to the greatest possible degree, use the existing data stored in various administrative sources. We have at our disposal high quality information infrastructure consisting of various administrative registers and databases, such as Population Register, Business Register, Register of Territorial Units, Tax Register, Employment Register, etc. These registers contain unique identifiers that allow to link data with other registers. This enables us to use their contents for different purposes, which is one of the strategic aims of the Statistical Office. Good results of this approach in previous censuses, as well as further development of administrative registers were

1 Prepared by Milan Katic.

the basis for the *strategic decision* that future censuses will be to a great extent register-based.

3. This strategy has implications for the role and function of IT. It is no longer focused merely on technical operations, such as data transfer from paper questionnaires to electronic media, editing, tabulation, etc., but on multipurpose use and integration of different data sources. However, the successful implementation of this task requires thorough preparation, as well as good management and organisation of work.

4. The paper describes in more detail the key phases of implementing the 2000 censuses in which the role of IT is especially important, i.e. Census preparation, data entry, editing and archiving, data aggregation, and output and dissemination.

5. The main IT supporting functions in individual phases of the Census are the following:

- (i) Census preparation
 - collection of the personal and other relevant data from the existing sources;
 - preparation of the supporting documentation (charts and maps of census districts, lists of houses and persons);
 - pre-printing personal data on the questionnaire.
- (ii) data entry, editing and archiving
 - functional integration of individual phases;
 - scanning and archiving of the questionnaires;
 - OCR (Optical Character Recognition);
 - data editing with direct access to the scanned questionnaires;
 - design and implementation of the microdata base;
 - archiving of the microdata.
- (iii) data aggregation and analysis
 - design and implementation of the census macrodata base (including aggregation on different levels);
 - integration of aggregates from other sources;
 - support to analytical functions (especially OLAP);
- (iv) output and dissemination
 - tabulation;
 - publishing the results (including in electronic format);
 - Internet and Intranet dissemination.

Census preparation phase

6. Extensive and thorough preparations are needed. The various organisational and technological tasks can only begin after the problems regarding contents and methodology are solved, and the general deadlines for the implementation are set (a special legal act is adopted).

7. One especially important task is the specification of relevant administrative sources and data to be collected from these sources. These are:

- personal data that will be pre-printed on census sheets;
- data from the Register of Territorial Units, on the basis of which detailed maps of census districts and the list of houses in the census districts will be prepared.

8. The IT support for setting up the pre-census databases in which personal data and data on census districts will be integrated, and the preparation of the above-mentioned documentation will be partly outsourced.

Data Entry, Editing and Archiving

9. The general strategy of data entry includes scanning census questionnaires and OCR technology of data capture. Scanned documents (images) will be stored in the *documentation database* and archived. The digitized records obtained using the OCR technology will be stored in the *input database*. In later phases of data editing, the interactively accessible database of scanned documents will be used as the basis for correcting the data. We will evaluate the possibilities to outsource the scanning and data capture using OCR for the population census, since we still have some time before taking the decision.

10. In these phases, IT plays a universal role. *Integration of scanning and OCR processes* is especially important as well as interactive use of the documentation database in the editing phase. The data editing (control and correction) will be performed by the Statistical Office using our own applications. During pilot censuses, the Blaise system for editing was successfully tested. Blaise will also be used for the Census of Agricultural Holdings in June 2000.

Aggregation and Analyses

11. Aggregation will be performed on the basis of clean data in the microdata base and the macrodata base will be set up. The level of aggregation is the statistical district as the territorial unit that enables aggregation into larger territorial units such as settlements, municipalities, regions and the entire country.

12. The macro database of aggregates will contain, in addition to census aggregates, other aggregate data from administrative and statistical sources (e.g. 1998 and 1991 Censuses). Thus, a single database will be set up for various analytical purposes and for post-census research of socio-economic phenomena.

13. In this phase the IT fulfils an *integrative* role:

- the integration of census aggregates with data from other sources;
- supporting on-line analytical functions with corresponding OLAP tools.

Output and Dissemination

14. In this phase, the role of IT has changed most compared to previous censuses. The census results will be made available via Internet and disseminated in electronic form in addition to printed publications. This opens up possibilities for a large-scale use of data for analyses and processing. It is planned to implement a tabulation system, which will produce all census output tables, as well as publication and dissemination of census results both in paper form and through electronic media. It is also planned to prepare an Atlas of Slovenia where the main census data will be presented on maps and charts. The highest level of aggregate data will be accessible via Internet.