UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.74/L.34
14 July 1982

ENGLISH ONLY

FOURTH UNITED NATIONS CONFERENCE ON THE
STANDARDIZATION OF GEOGRAPHICAL NAMES
Geneva, 24 August-14 September 1982
Item 9 of the provisional agenda*

AUTOMATED DATA PROCESSING

Automation of Canada's National Toponymic Data Base: 1977 to 1982**

Paper presented by Canada

* E/CONF.74/1.

** Prepared by Mr. M. R. Munro, Head of the Toponymy Section, Surveys and Mapping Branch, Department of Energy, Mines and Resources.

At the Third United Nations Conference on the Standardization of Geographical Names, held in 1977, delegates were informed that the Surveys and Mapping Branch, Department of Energy, Mines and Resources, Canada, had begun to investigate how best to automate the geographical names data used in Surveys and Mapping Branch products. This process has led to a significant number of developments as outlined below.

Decision to Automate

The methods for the handling of geographical names by the Surveys and Mapping Branch for its own maps and publications, for other mapping agencies and for other users, were examined.

It was found that with large amounts of clerical manpower required for the storage, retrieval and updating of these records, especially in a time of severe restrictions on manpower usage, an alternative was needed.

Storage space in itself was an identifiable problem in terms of the volume of records, filing cabinets and floor space.

Gazetteers of the Gazetteer of Canada series and their Supplements were produced in the 1970's by sending the card records to typesetting companies. Not only did this involve a large amount of clerical work, but the security of the records was threatened. If they were lost or damaged, years of work would be required to replace them. Interestingly, these service companies used computers to input the names data.

Having looked at the existing methods for the handling of the information in this toponymic data base maintained by the Secretariat of the Canadian Permanent Committee on Geographical Names and assessing the liabilities of those methods, it was felt that automation was well-suited to cope with many of these problems. The records were already in a form that could be easily adapted to a computer; clerical manpower requirements for updating, retrieval and storage could be greatly reduced; storage space would be maximized; security could be ensured; and the ability to better meet user needs for maps, publications and lists, were all points favouring automation.

Automation

Having looked into the existing methods for the handling of toponymic information and having decided that automation was the most practical way of meeting the Branch's needs, the next step was to evaluate how to go about automating.

An evaluation of the basic toponymic needs of the Surveys and Mapping Branch was made in-house in 1977, after which a consulting firm was contracted to recommend how best to apply computer technology. The consultant recommended a dedicated mini-computer with custom designed software, but since there were a number of good, reasonably inexpensive commercial software packages on the market, some of these were looked at. One of these, DATABOSS/2 and a RSTS/E operating system was selected because it could operate on a Branch PDP 11/45 mini-computer, because it could be easily adapted to the Branch's needs with a minimum of programming, because it operates in the basic-plus programming language and because word-processing functions can be performed on it.

DATABOSS/2 is essentially a single-file system with a maximum record length of 512 character-spaces, but this was found to be inadequate for the needs of the Branch and the single file was then divided into seven files. Rather than discussing each of these files, their fields of information and their respective character-space lengths, a brief overview of the information contained in the automated data base and how it can be used is given.(1)

a) Each entry shows the following information where applicable:

1. Feature name - each name is assigned a unique 5-letter key for accessing information from other files;

2. Generic - each name's generic is identified by a four digit number code.

3. Province or Territory - a two digit code is used to identify the province or territory in which the named feature is located, (undersea features are also assigned a separate "province" code in order to identify that they are in waters within areas of Canada's interest).

4. Decision date and status - each name is assigned a status code. (e.g. a name with an A1 status is a currently approved name, while a name with a B4 status is a name formerly approved but is not approved at present, having been replaced by another name), and the actual decision date, (e.g. indicating the latest ruling on the status of a name).

5. Location - the location of a named feature is indicated by a "province" code, by latitude, by longitude, by cadastral information (e.g. county, township or section) and by a location narrative (e.g. "30 km. west of Kentville").

6. Map - the occurrence of geographical names on federally - produced maps or charts is indicated, (e.g. 105H14, a series number of map of the National Topographic System).

7. Origin - any information available on a name's origin and meaning is entered in a narrative format.

(1) Thompson, John J.S. Automation of The National Toponymic Data Base, CANOMA, Vol. 6 No. 1, pp. 7-11, Surveys and Mapping Branch, Department of Energy, Mines and Resources, Ottawa, Canada, 1980.

8. Cross-Reference - any cross-referenced or variant name is included.

b) A variety of uses can be made of the system.

Data entry can be performed by use of a custom-designed program which stores this new information on temporary files. Once approved for entry into the master file, another custom-designed program, the "bulk entry program" merges the information stored in the temporary files into the master file.

Information in the master file or on a temporary file can be changed, deleted or displayed by authorized operators.

A "report generator" program designed in-house allows an operator to obtain specified lists (e.g., all names in Nova Scotia beginning with the letter "B"; or, all approved names on a specified map sheet; or, all names within a quadrant defined by latitude and longitude).

c) Publishing of the data base (e.g. gazetteers) no longer requires clerical staff to remove cards from file drawers to be sent to photo-typesetting companies. Instead, formatted magnetic tapes can be sent for use by automated typesetting machines or laser printers. This is more efficient, less costly, a better use of manpower and safer since invaluable records do not have to leave the premises.

d) Security of the records has been greatly enhanced not only because the originals do not have to be sent off the premises for printing purposes, but because they are in a computer which denies access to anyone without the authorized passwords, thus preventing vandalism or human error. Even if vandalism, a natural castastrophe (i.e. fire, earthquake, etc.) or human error were to destroy the computer and the data base, a back-up system is maintained elsewhere.

e) The Cartography and Toponymy Division of the Surveys and Mapping Branch is now using with the RSTS/E operating system, the DEC ED2 word processing function.

The system is still operating with the same software package and custom programming based on the original feasibility study. However, after more than two years of use, the system underwent a post-implementation evaluation in 1981.

The post-implementation study, conducted by contract has made the following recommendations:

- that the data base be restructured into 13 "province" files (i.e. one file for each of the 10 provinces, one file for each of the 2 territories and one file for undersea features within areas of Canada's interest);

- that via a video terminal, all information in the seven files of the data base for any individual record can be accessed, portrayed, changed or deleted as a single unit;

- that a program be written specifically for the production of names lists for new or revision mapping (i.e. a single one-line command to generate a list of the names to be shown on a particular map). The production of these lists is a highly important use of the NTDB and such a program would facilitate their production;

- that a program be written specifically for the production of formatted gazetteer reports (printouts) and formatted gazetteer magnetic tapes for printing purposes;

- that a function be made available to authorized users with terminals to access the NTDB information through on-line computer communications.

Hardware:

The NTDB is operated on a DEC PDP 11/45 computer with 128K words of memory, assisted by a TE/10 bpi tape drive, four RP06 disc drives (two have just recently been purchased) each with 176 Megabytes, two DZ-11 8-line multiplexors, one DC-11 communications interface (dial-up), one HP 2631A high-speed printer, one letter-quality printer (Multiwriter), one Volker-Craig 414 video display terminal and seven Volker-Craig 404 video display terminals.

Content of the NTDB

There were an estimated 350 000 card records to be entered into the automated NTDB, of which over 90% have now been entered. This loading should be completed early in the fall of 1982.

Since there are expected to be approximately 25 000 new records to be added each year, (the average space allotment per record is 800 character spaces for data items and 100 character spaces for management information), the disc storage has been increased to accommodate 600 000 toponymic records, (e.g. 540 million character spaces).

Conclusions:

The automated NTDB, when fully operational, will have taken almost five years to completely design, program and load from the date of the first feasibility study. It will provide a wide range of services to the Branch, other federal departments and to other users. Many of these services are now in use but in addition to those already discussed, the NTDB is going to play an important role in the computer assisted cartography of the future and in interfaces with other computer facilities in the many disciplines using geographical names.

It should be noted that apart from it being an effective way of keeping large amounts of data, thus reducing person-year requirements and associated costs to a minimum, and providing a quick way of responding to enquiries, the NTDB is a very "practical" data base. It is used in the production milieu, and can provide easy-to-understand material for clerks, researchers

and academics alike, and does not deal with subjective material (i.e. some administrative data bases rely heavily on the information provided by individuals who may or inadvertently, may not give accurate data), but rather with a controlled and factual set of data reflecting a long-established, conventional, manually maintained data base.

The next report about the automated National Toponymic Data Base to be presented at the 1987 Conference, will show the effectiveness of this system in terms of its role in the rapidly developing computer technology being applied to cartography.
