



Energy, Mines and
Resources Canada

Énergie, Mines et
Ressources Canada

CANOMA

Vol. 12 No. 1

July/juillet 1986

AUTOMATION OF GEOGRAPHICAL NAMES

AUTOMATISATION DES NOMS GÉOGRAPHIQUES

SECOND SYMPOSIUM
OTTAWA
MAY 31, 1985

DEUXIÈME COLLOQUE
OTTAWA
LE 31 MAI 1985

Canada

Canadian Permanent Committee on Geographical Names
Comité permanent canadien des noms géographiques

COVER:

Displayed are two computer plots, generated separately on plastic and superimposed. The geographical names illustrate a selection of character fonts available on the computer. The black overlay was created in block letters, using a thicker pen. Both layers were plotted on a CALCOMP 1051 plotter using a graphics package called DISSPLA.

(Courtesy: Computer Science Centre, EMR)

COUVERTURE:

Deux tracés furent reproduits séparément sur plastique et ensuite superposés. Les noms géographiques démontrent le choix de caractères d'imprimerie que l'ordinateur peut produire. Le calque de superposition noir, en lettres moulées, fut produit avec une plume moins pointue. Les deux tracés furent produits par un traceur CALCOMP 1051 en utilisant un système d'écriture du nom de DISSPLA.

(Gracieuseté: Centre d'Informatique, ÉMR)

Communications concerning this publication or geographical names in general should be sent to:

Secretariat
Canadian Permanent Committee on Geographical Names
6th Floor, 615 Booth Street
Ottawa, Ontario, K1A 0E9

Toute information concernant cette publication ou les noms géographiques en général devrait être envoyée au:

Secrétariat
Comité permanent canadien des noms géographiques
6^e étage, 615, rue Booth
Ottawa, Ontario, K1A 0E9

Compilation and layout/Compilation et disposition typographique:

Denise Patry

With assistance from/avec l'aide de:

Helen Kerfoot
Alan Rayburn
René Leduc
Kathleen O'Brien
Ken Lightfoot

©Minister of Supply and Services Canada
1986

Catalogue No. M85-12/12-1

ISSN 0319-5228

©Ministre des Approvisionnement et Services
Canada 1986

N^o de catalogue M85-12/12-1

ISSN 0319-5228

CANOMA

Vol. 12 No. 1

July/juillet 1986

AUTOMATION OF
GEOGRAPHICAL NAMES

AUTOMATISATION DES
NOMS GÉOGRAPHIQUES

sey
BRONTO
River • Fru
Pierre • SASKE
NGNIRTUNG • Renfo
REGINA • Joachin
Anchorage • Tinhori
and • Pointe à Péco
PELICAN PORT
CHARLOTT
OWS • F
D C

PROCEEDINGS OF THE
SECOND SYMPOSIUM

COMPTE RENDU DU
DEUXIÈME COLLOQUE

OTTAWA
MAY 31, 1985

OTTAWA
LE 31 MAI 1985

Published by the Surveys and
Mapping Branch, Energy, Mines and
Resources, Canada

Publié par la Direction des Levés et
de la Cartographie, Énergie, Mines
et Ressources, Canada

CONTENTS / SOMMAIRE

	<u>Page</u>
<u>Papers presented/Communications présentées:</u>	
Excerpts from the opening remarks of Jean-Paul Drolet	1
The National Toponymic Data Base - a progress report	J.J.S. Thompson and M.R. Munro 2
Ontario's Toponymic Data Base: summary of user needs survey	Jeffrey R. Ball 9
Toward a new system [Translation]	Commission de toponymie du Québec 16
Automated location of toponyms in their respective municipalities [Translation]	Clément Nolette 19
Geographic Names Information System (GNIS): philosophy and function	Roger L. Payne 21
<u>Brief reports/Exposés:</u>	
• TERMIUM III	John Carey 25
• Canadian Hydrographic Service report on the automation of undersea feature names	Colin Bromfield 26
• Statistics Canada: activities using geographical names	Henry A. Puderer Joel Yan 26
Panel discussion and commentary	-- 31
<u>Communications présentées/Papers presented:</u>	
Extraits du discours d'ouverture de Jean-Paul Drolet	37
La base nationale de données toponymiques - compte rendu des travaux [Traduction]	J.J.S. Thompson et M.R. Munro 38
Base de données toponymiques de l'Ontario: compte rendu d'une étude des besoins des utilisateurs [Traduction]	Jeffrey R. Ball 45
Vers un nouveau système	Commission de toponymie du Québec 53
Situation automatique d'un toponyme dans sa municipalité	Clément Nolette 56
"Geographic Names Information System (GNIS)" - principes et fonctions [Traduction]	Roger L. Payne 58
<u>Exposés/Brief reports:</u>	
• TERMIUM III	John Carey 62
• Rapport du Service hydrographique du Canada sur l'automatisation des noms d'entités sous-marines	Colin Bromfield 64
• Statistique Canada: activités portant sur les noms géographiques	Henry A. Puderer Joel Yan 64
Échange en panel et commentaire	-- 69

EXCERPTS FROM THE OPENING REMARKS

OF

JEAN-PAUL DROLET*

It gives me considerable pleasure to welcome you to the Second Symposium on the Automation of Geographical Names. The large attendance here today is a reflection of the strong interest taken by the various names boards and mapping offices in Canada and the United States in achieving practical and reliable filing systems for geographical name information and in improving the capabilities of exchanging information through automated means.

In convening this symposium I believe the Canadian Permanent Committee on Geographical Names is performing some of the important functions for which it was established in 1897: the task of standardizing geographical names throughout Canada and the role of bringing together personnel from the various provincial and federal names authorities to resolve problems of common concern.

First, I would like to thank those who have put together presentations for this symposium: John Thompson of the federal Surveys and Mapping Branch; Jean Poirier of the Commission de toponymie du Québec; Jeffrey Ball of the Ontario Ministry of Natural Resources; Clément Nolette of the Département de l'Informatique of Université Laval, and Roger Payne of the Domestic Names Section of the United States Geographic Board. We have also with us the following people who have agreed to serve with the speakers on the special panel this afternoon: Richard Darley of National Geographic Magazine, Washington; Donald Orth of the United States Geographic Board; and Gerald Holm of the Manitoba Department of Natural Resources. Also in attendance at our meeting is Randall Detro, from Nicholls State University in Thibodaux, Louisiana. He is the director of the United States Place-name Survey.

The program is designed to have five formal presentations followed by a panel discussion, after which the audience may participate in a question-and-answer session. I want the program to be flexible enough, to allow for contributions from others who are involved in the development of automated systems relating to names. I understand Colin Bromfield of the Canadian Hydrographic Service, John Carey of the Translation Bureau and Joel Yan of Statistics Canada have put together some material

on automated programs in their offices, and are prepared to make brief presentations.

The First Symposium on the Automation of Geographical Names was held in Ottawa almost exactly five years ago. We had some excellent presentations on the theoretical and developmental aspects of automated systems, on the implementation of systems and the requirements of users for reliable and unambiguous geographical name information.

At the 1980 symposium I raised a number of fundamental questions about automating geographical name information: those questions along with several questions brought up during the symposium bear repeating again this morning.

1. How can modern computer technology be used to record geographical name information?
2. How can computer technology make names information more readily available to users?
3. How do we ensure compatibility between different systems to allow the efficient exchange of essential information?
4. What are the content limits of a toponymic data base?
5. How can the quality and availability of names information be improved at a time when spending restraints discourage increasing the number of people recording and handling names information?
6. How can names information be protected from change or alteration once it has been authorized by the appropriate names authority?
7. What are users' expectations?
8. How far should users' be expected to repay the costs of system development and management?
9. At what point should traditional hard copy records be discontinued and reliance then placed totally on the automated systems?

* Dr. Jean-Paul Drolet, Chairman, Canadian Permanent Committee on Geographical Names.

These are just a selection of questions requiring answers.

★★★★★★★★★★

THE NATIONAL TOPONYMIC DATA BASE - A PROGRESS REPORT

J.J.S. Thompson* / M.R. Munro**

ABSTRACT. The National Toponymic Data Base (NTDB) is an integrated digital and analogue data base of geographic names information maintained by Energy, Mines and Resources, Canada. Significant changes have taken place in the structure of the NTDB since it was first automated in 1978, specifically, the number of files has been reduced to facilitate system transactions. A review is presently underway to evaluate future NTDB hardware and software requirements, in order to respond to increased demand, a growing volume of names records, and a commitment to a more distributed National Toponymic Data Base.

INTRODUCTION

The National Toponymic Data Base (NTDB) is an integrated digital and analogue data base of geographic names information, including toponyms approved by the Canadian Permanent Committee on Geographical Names (CPCGN). The Toponymy Section, Surveys and Mapping Branch, is responsible for the operation and maintenance of the data base, and through the offices of the Secretariat of the CPCGN, ensures the availability of toponymic information to the Canadian federal government and its agencies, and to the general public.

Some time ago, it became evident that continuing growth in the number of traditional, analogue toponymic records, and an ever-increasing demand for more efficient names information acquisition, storage and retrieval would necessitate an evaluation of the traditional, labour-intensive, toponymic information handling methods used by the Toponymy Section to maintain and operate the geographic names records system.

INITIAL AUTOMATION

In 1977, the Geographical Services Division undertook a survey of the services provided by the Toponymy Section and of the quality of the toponymic records in use at that time - specifically, carded information and documents in paper files. The survey indicated that due to the large size of the data base, (estimated in 1977 at 350 000 names recorded in primary card files and 650 000 records in secondary folder files), and a 7% annual rate of additions and revisions, the toponymic system required the benefit of more modern technology in order to remain responsive to user requirements. To provide for the efficient storage and retrieval of this information, and in order to make the toponymic data base more compatible with various automated aspects of cartography, it was decided to computerize the toponymic files.

Following the 1977 evaluation, a national computer software consulting firm was contracted to provide recommendations for the automation of the NTDB. The report

submitted by the contractor recommended a customized data base management software package operating on a mini-computer. However, after careful review, Division management felt that due to the various software packages with significant potential that were available commercially, several of these should be evaluated in order to determine their suitability for use in the toponymic data base. Ultimately the software package decided upon was DATABOSS/2. The system, written in the Basic-Plus programming language, can be used in either on-line, or batch mode, takes input from any ASCII file, and is run on DEC PDP mini-computers and the RSTS/E operating system. The DEC PDP and RSTS/E compatibility were especially important, as the Division had access to an existing DEC PDP 11/45 computer at the time.

As a pilot project, a decision was made to produce a new edition of the *Gazetteer of Canada: Northwest Territories*, using the automated system. The toponymic card information was manually transcribed onto computer coding forms and then sent to a word processing firm for the preparation of a magnetic tape. At the same time, IAS Computers Ltd. was contracted to supply and install the DATABOSS/2 software package and the RSTS/E operating system on a Branch PDP 11/45 computer. The system was brought on-line in accordance with detailed storage and retrieval specifications and the tape of the Northwest Territories toponymic information was then loaded onto the system. The information was edited on-line, and a final tape was produced from the new system. The *Gazetteer of Canada: Northwest Territories, 1980*, became the first gazetteer output of the automated National Toponymic Data Base. Since that time, five other gazetteers have been produced for this series, including those for Manitoba, Yukon, Newfoundland and British Columbia. Work on the Saskatchewan volume is now in progress, with publication scheduled for the summer of 1985. The NTDB is continually updated and corrected to reflect the latest provincial names decisions.

The DATABOSS/2 package is characterized by a single file system, which while suitable for a large number of the records, places a number of constraints on the toponymic system, not the least of which is record

* John J.S. Thompson, Assistant Director, Geographic Mapping, Geographical Services Division, Surveys and Mapping Branch, EMR.

** M.R. Munro, Head, Toponymy Section, Geographic Mapping, Geographical Services Division, Surveys and Mapping Branch, EMR.

length. In order to optimize the commercial system, in-house programming was used to transform DATABOSS/2 into a multi-file system of seven data base files. These files consisted of record sets called: Gazette Data Base File (which included a unique identifier, fields for the feature's name, the generic, the province, the decision date, the status, latitude, longitude, fields for gazetteer production and type sizing); Map Sheet Data Base File; Cross Reference Data Base File; three data bases dealing with the name's location; and the Origin Narrative Data Base File. (Figure 1).

As a result of this work, all of the desirable features that had been present in the original automated system were maintained, but from the user's point of view record manipulation was greatly enhanced by the revised programming. In addition to this, several new functions were created which facilitated use of the system. The new programs optimized much of the remaining system capability (based on the seven data base file structure). Unfortunately, these new programs left little room for any further system development, and despite various improvements, system users continued to feel

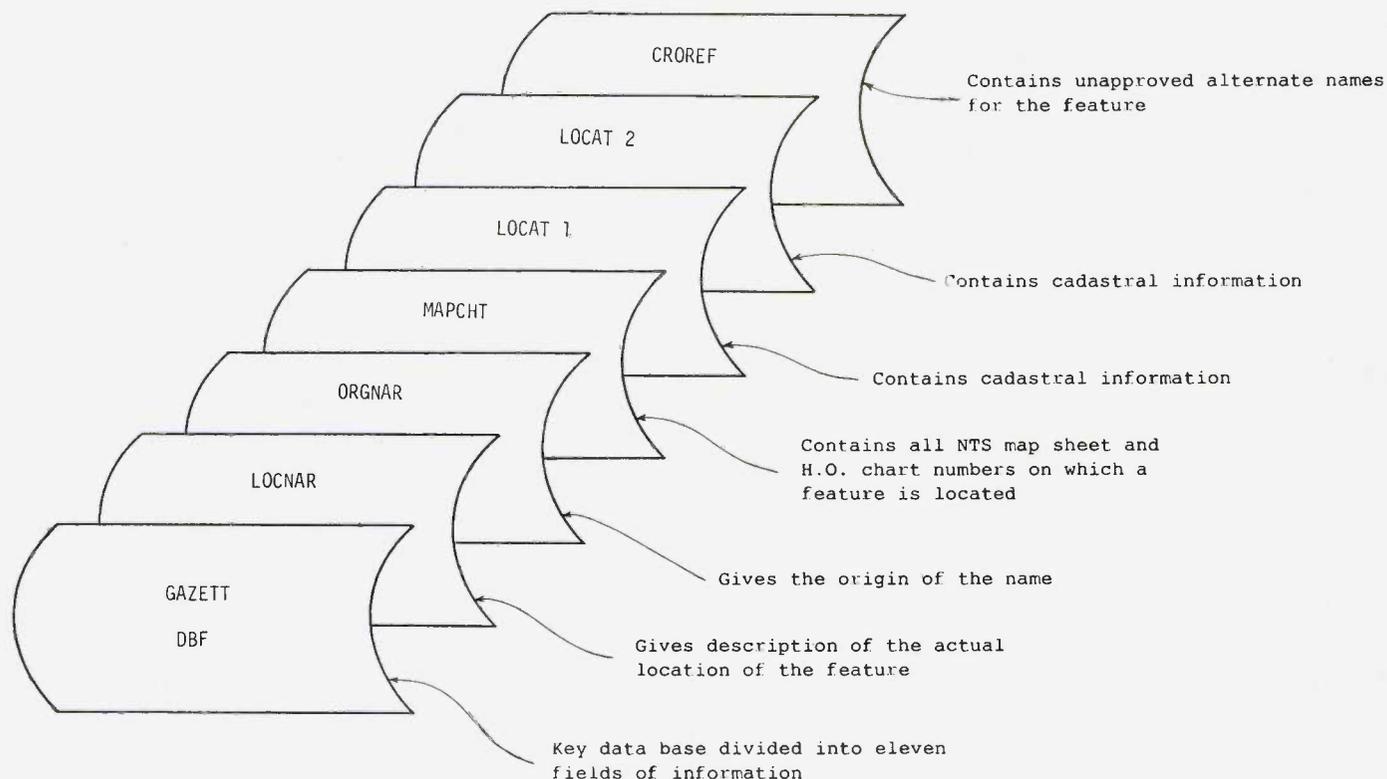


Figure 1 Original National Toponymic Data Base showing seven data base structures

Operations available to users of the system involved the creation and display of data base files and the insertion, deletion, display and editing of data base records. In addition to this, three programs were developed for the system, two of which were specifically designed for data entry, called the Interactive Data Entry Program, and the Bulk Entry Program, and a program called the Report Generator.

In general, it can be said that the initial attempts at automation created a system that was more functional than the original analogue system in that it allowed more flexibility for names information processing than was ever before possible. However, it can also be said that many of its operations were slower and more cumbersome than had been expected from an automated system. It was also clear that due to hardware and software constraints, the performance would continue to degrade as the data base grew. As a result, some system reprogramming was carried out in order to try to minimize the constraints.

hampered by the seven file structure, especially as users were required to access each of the seven related files individually in order to perform record operations.

PRESENT CONFIGURATION

As a result of the continued data base growth, and increasing demand for quicker system response, a new system configuration has been implemented that has resulted in two significant system architecture changes:

- The consolidation of the original seven files (Figure 2) into four expanded files, specifically: Text, Map Chart, Narrative, and Gazette files, (Figure 3).
- the construction of 14 required sub-files (10 for the provinces, 2 for the territories, and 1 for under-sea features, and 1 for international waters).

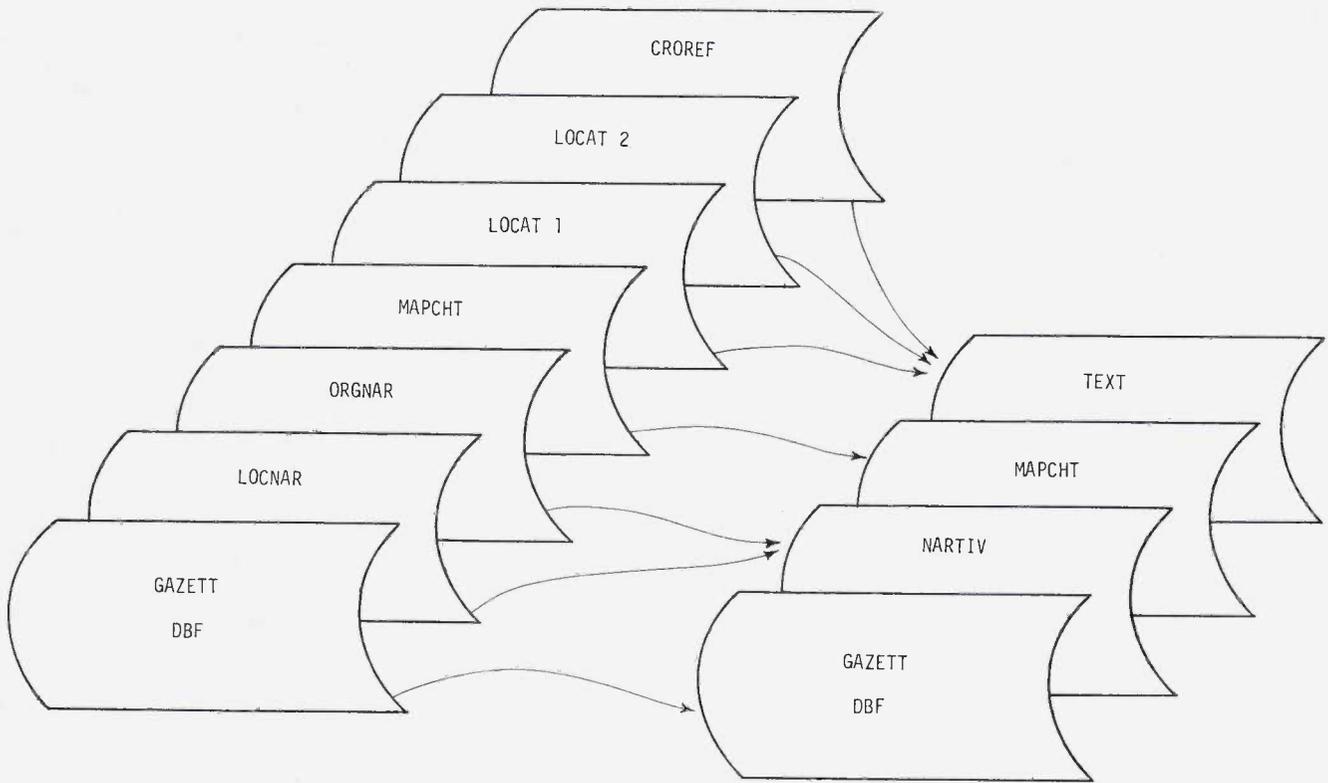
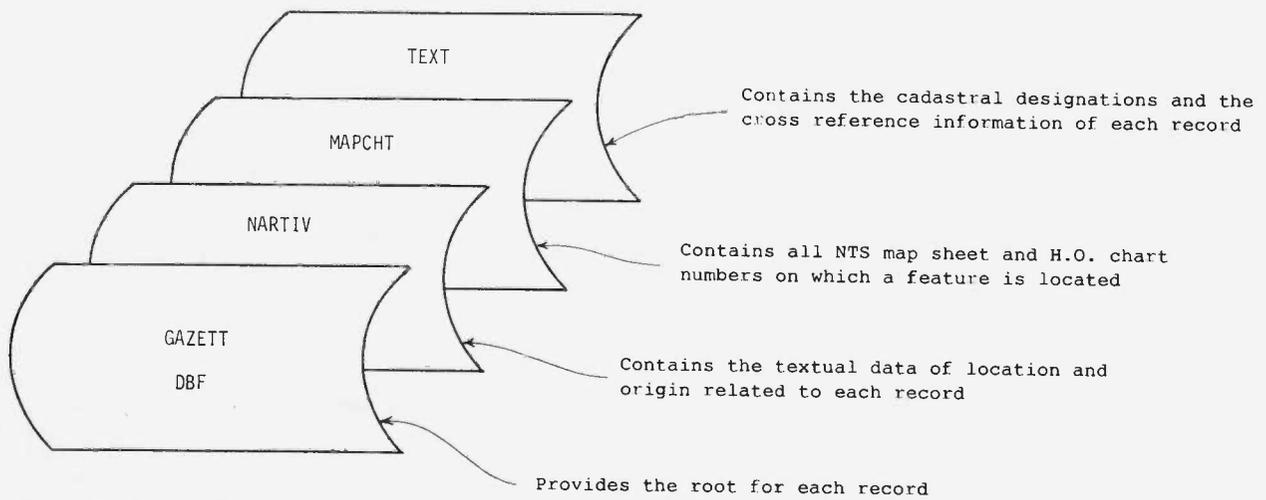


Figure 2 Consolidation of seven original NTDB files into four revised files



(Each file is in turn sub-divided into 14 regional "data bases")

Figure 3 Present National Toponymic Data Base structure

This re-adjustment of the data base file structure now provides more rapid access to the data and ensures that the system file structure will be able to more easily accommodate the projected number of feature names. System management has also been enhanced, allowing easier maintenance and record audit. The existence of the four separate data base files is transparent to the user through the custom software, allowing access to each record as a single unit. This approach permits much quicker retrieval, editing, and addition or deletion of any given record.

In addition to these changes, a number of new programs have been written which while providing all of the functions available with the original system and the subsequent re-programming, have streamlined the mechanisms involved and have provided some totally new and very useful services.

CUSTOM PROGRAMS COMPLETED TO DATE INCLUDE

a) Source Data Entry

This program, like the original Interactive Data Entry Program, is specifically designed to provide a means of entering large volumes of new records into the relevant regional data base. The user can use this program to enter records into source files and perform delete, edit, and print functions on the data being entered. The data base manager can then authorize close and merge operations to enter the information into the main data base via a mass insert operation.

b) On-line Data Base Transactions

With this program, transactions can be performed "on-line" with the material in the main data base. This includes delete and update as well as provisions for recovery of transactions in the event of data loss. Like the Source Data Entry Program, the data base manager authorizes the merging of these transactions.

c) Nameslist Reports

This program was designed specifically for the staff of the Toponymy Section who prepare nameslists for the ordering of type by Branch cartographers. There were a number of operations that had to be included in this program. The tasks performed are: creation of a generic set file; initiation of a nameslist report; the printing of the nameslist report; the creation of a nameslist format control file which allows staff members to interactively define and edit the toponymic data required for inclusion in the report; and the production of the revised report and the saving of this final report for inclusion in a magnetic tape for future regeneration when the respective map sheet is being revised. The fact that these nameslists are stored digitally means that a great saving in time will result in future sheet revisions. (Figure 4).

d) Gazetteer Production

The Toponymy Section is responsible for the production of the Gazetteer of Canada series, and two programs were established for this purpose. The first of these is called the Gazetteer Report Program, which allows the data base manager to generate a formatted print-out of material as it is to be shown in a specific gazetteer. This print-out is checked against source material and edits made to the main data base. In terms of assisting this work, small programming changes have been made to

the system so that the number of occurrences of specific items (i.e. generics), can be counted, or the feature names of a certain type can be printed-out. (The generics referred to are the generic elements of toponyms). The second program written in this series is the program used in the creation of a gazetteer production tape. This program generates an encoded magnetic tape which can be sent to a typesetting company to produce the master copy of the gazetteer needed by the printers. This tape includes all the currently approved geographical names as well as any formerly approved names that have either been changed or rescinded.

e) Fieldwork and Decision List

This program has been designed to allow the creation of a temporary file in which toponyms can be inserted, edited, deleted, printed and, with the authorization of the data base manager, transferred from the temporary file to the decision list file. The data base manager can then authorize the merging of this file with the main data base.

f) General Information Retrievals

As the name indicates, this program allows the user to call up any toponym or group of toponyms in the main data base for printing or viewing purposes. In addition to this, retrievals can also be made using an "AND" or an "OR" logic, allowing the user to define specific fields. For example, the system can be queried to display all of the features in New Brunswick that contain the generic "river" AND contain the status code "A1" (approved toponym). Further, the system could be asked to display all of the generics "river" OR all of the features with the generic "lake" in a specific region. The area of query can be further delimited by various different parameters, including the provision of a geographic window defined by latitude and longitude.

g) Data Base Management Programming

Several programs have been written which make the management of the system more efficient.

Among these are:

- 1) a program written to assign NTS map sheet numbers to any records having latitude and longitude coordinates within that map sheet area;
- 2) a short "skeleton" program for each customization of existing programs;
- 3) programming that will allow the data base manager to run jobs in batch mode at specified times;
- 4) programming to change job priorities while different jobs are being run concurrently;
- 5) programming to allow the interruption of programs being printed so that they can be restarted without having to repeat work already done;
- 6) programming so that searches can be made in a detached mode;
- 7) programming so that jobs to be printed can be queued, thus allowing work to be done without operator intervention;

```
Map Sheet / Carte: 62 K/12 SPY HILL SASKATCHEWAN MANITOBA
Project # / Projet #: 02-1916
Sheet Title Block / Titre de la Feuille:

SPY HILL
SASKATCHEWAN MANITOBA

----- PROVINCIAL BOUNDARIES / FRONTIERES PROVINCIALES

Saskatchewan F-3
Manitoba F-3

No. Name / Nom Location
-----
----- PLACE NAMES / NOMS DE LIEU

1 Yarbo A-3
2 Cutarm B-3
3 Gerald C-3
4 Spy Hill D-2
5 Hazel Cliffe A-1
6 Tantallon B-C-1
7 Weiby F-1

----- LAKES, BAYS ETC. / LACS, BAIES ETC.

8 Miller Lake D-2
9 Ferrin Lakes E-2
```

Figure 4 Sample automated names list output - NTDB

8) programming to provide an ability to close and submit transactions remotely.

SECURITY

The data base continues to be safeguarded by the use of user accounts, protection codes and passwords. This allows the data base manager to control access to the data base, and while this element has been present since the establishment of the original system, the re-programming has made control much tighter. The required authorization of the data base manager significantly reduces the chance of unapproved changes or deletions

h) Generic Description File

This program is integral to many others used in the system in that this file can be linked with a number of the other programs and in particular with the Gazetteer Report Program.

to the NTDB.

Back-ups are made on both a daily and weekly basis, with a regular back-up of the entire system performed every two weeks. These back-ups are stored in a secure area in the event of any catastrophic failure or vandalism.

NTDB SYSTEM HARDWARE

The hardware used with the NTDB has changed little since the system was originally implemented. The main differences have been an increase in the amount of data storage space and the acquisition of a high-speed printer. Most of the original terminals have now been replaced by newer equipment, but while newer, the functions remain essentially the same. The components presently comprising this system are: (Figure 5)

- 1) a DEC PDP 11/45 computer with 256 KB main memory;
- 2) fifteen disk packs;
- 3) one TE16 (800/160 BPI, 45 IPS) dual density tape drive;

- 4) five RP06 disk drives, each 176 Megabytes;
- 5) two DZ-11 multiplexors (providing up to 16 terminal capability including dial-up support);
- 6) one General Electric 3400 high-speed printer;
- 7) three Volker-Craig 404 terminals;
- 8) one Volker-Craig 414 Word Processing terminal;
- 9) two Cybernex XLA100 terminals;
- 10) five Cybernex XLA84F terminals;
- 11) one Digital Equipment Corporation VT100;
- 12) one Multi-Writer high-quality printer;
- 13) one LA34 hardcopy terminal (console)

THE FUTURE

A review is presently underway to determine the direction of future developments in the automation

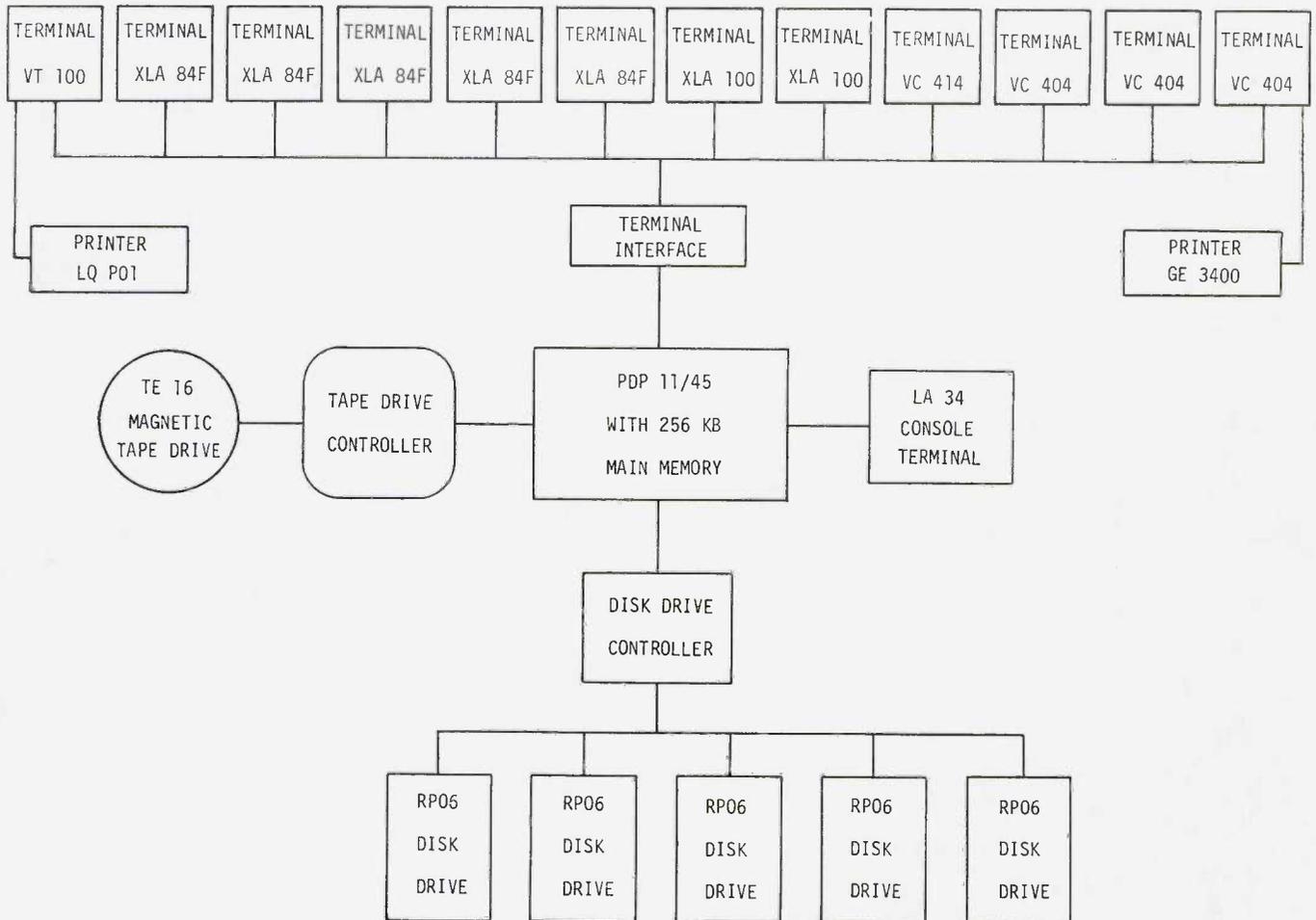


Figure 5 NTDB system hardware

of the National Toponymic Data Base. It is proposed that by 1987 a more modern computer will be required by Surveys and Mapping Branch for the NTDB. The current computer is archaic by industry standards. It is extremely slow, there are frequent breakdowns and its use is limited to a relatively small number of users. If the NTDB is to play a major role in the timely, increasingly computerized production of maps, this problem will have to be resolved.

The requirement for new software is also being addressed. The existing software has been customized and re-programmed to the point where it has filled all the space available for programming. While the initial acquisition of the DATABOSS/2 software package eight years ago provided a low-cost, relatively straightforward introduction to the automation of the NTDB, it is becoming apparent that the combination of a strong demand for digital toponymic information, and decreasing human resources available for the manual processing of geographic names queries, dictate the need for significant system development.

A principal focus of any further development of the NTDB will be an effort to ensure maximum accessibility of the data base to as many users as possible. It is hoped that over the next several years the NTDB will become, increasingly, a distributed data base - one to which provincial names authorities have ready access, on at least a "read-only" basis. We are committed to increased cooperation between NTDB staff, the Secretariat of the CPCGN and provincial names authorities regarding automated gazetteer production, and the mutual exchange of digital toponymic information.

Increasingly, the Toponymy Section is receiving requests from provincial and federal agencies, and from the private sector, for computerized toponymic listings and digital file information. Towards this end, standard format tapes are being developed that will facilitate response to these user demands.

As stated previously, the rapidly deteriorating condition of our current computer host has necessitated a review of present and future NTDB system hardware and

software configuration requirements. Principal among these will be the need to facilitate rapid and accurate geographic names retrieval for digital and analogue mapping purposes, particularly maps of the National Topographic map series at various scales, and the International Map of the World series at the 1:1M scale.

Our ultimate objective in this respect is to provide timely, accurate digital toponymic information in appropriately sized and styled type, available to the cartographer, on demand, for any given map sheet.

CONCLUSION

Good progress continues to be made in the automation of the National Toponymic Data Base. Data base staff, many of whom have had no previous experience with computers and automation have responded in a positive and constructive manner to the program. Provincial gazetteers are being produced with significantly fewer human resources than was ever before possible, and the completion of each gazetteer substantially improves the currency of the data base.

The operating system has been refined to a point where many system transactions are transparent, and the system has, therefore, become increasingly user-friendly. A proven routine of check and cross-check, back-up and audit trail help to ensure system integrity and security. However, significantly increased demands for digital toponymic information, the continued growth of the data base, and the rapidly ageing computer hardware are now requiring the re-evaluation of a system hardware and software configuration needed to take the NTDB successfully through the next decade.

We are aware of the great potential that exists for the direct interface of the NTDB to outputs suitable for easy use in many of cartographic and geographic products, and look forward with enthusiasm to future NTDB development that will see the increased exchange of digital toponymic information with interested users, and the establishment of an easily accessible, distributed National Toponymic Data Base.



SYMPOSIUM PARTICIPANTS:

On the right: Jean-Paul Drolet (CPCGN Chairman)

On the left: John J.S. Thompson (Surveys and Mapping Branch, EMR)

ONTARIO'S TOPONYMIC DATA BASE: SUMMARY OF USER NEEDS SURVEY

Jeffrey R. Ball*

ABSTRACT. The province of Ontario's toponymic record is being redesigned to support the provincial geographical referencing system, and to provide an effective data base of approved and recorded geographical names. Stage one of the redesign is under way: a series of "graphic indexes" showing the Geographic Nomenclature of Ontario is being prepared in conjunction with the establishment of the Geographical Referencing System. Stage two will be devoted to automating the names information presently held in the manual system. In 1984 a user requirements survey was carried out as part of the automation phase. The results of the survey showed that users of the record generally require approved and recorded geographical names and the graphic extent of the named features and the places contained in the provincial map files. These findings are used as a basis for the recommendation of a strategy for automating the Ontario geographic names record. The strategy and its rationale are then examined.

INTRODUCTION

There are currently about 120,000 recorded names of geographical features and populated places in the province of Ontario. These names are recorded in a manual system, consisting of a card-index file and maps annotated with the names of the features. Approximately one-half of the names are approved and form an integral part of the cartographic data base for the province's geographical information system. The cards and maps constitute the only comprehensive record of geographical names in Ontario, and hold information that could be useful to other agencies.

The current system is antiquated and difficult to use, however. Access to, and retrieval of information is time-consuming, and responding to requests for names information is a laborious task. The base maps are fairly fragile and copies of this graphic information is not readily obtainable.

Ontario has decided to redesign the toponymic record to support the provincial geographical information system, and to improve the storage and retrieval of recorded names.

The first stage in upgrading the system has already begun. A series of graphic indexes showing the approved Geographic Nomenclature of Ontario is being prepared in conjunction with the progress of mapping in the Ontario Basic Mapping (OBM) program and the establishment of the Geographical Referencing System. The series is being compiled on the National Topographic Series (NTS) 1:50 000 map bases, which are available for a large part of the province and have most of the names information on them. The compilation is scheduled concurrently with OBM production to provide graphic indexes on stable film from which inexpensive white prints can be made.

Prints from the series are being distributed to the Ministry of Natural Resources Regional and District offices and are also available for provincial and federal agencies and other interested users.

The second stage of system redesign will be devoted to automating the geographic names records to provide faster access to information presently held in the manual system, and to facilitate revision of the data base.

In 1984 the province carried out a user requirements survey as part of the second stage of system improvement. This paper details the findings of that survey. Finally, a specific strategy for automating the Ontario geographic name records is proposed and its rationale examined.

SURVEY METHODOLOGY

The user requirements questionnaire was developed based on information obtained during a similar survey in 1980. This earlier survey was part of a feasibility study conducted by a consulting firm to determine if computer technology could assist in improving the delivery of nomenclature information to users of the provincial names record. The later survey was intended to update the information in the earlier survey before work was started to improve the system.

Very briefly, the later survey consisted of a six-part questionnaire (see Appendix A) used to identify:

- current and future requirements for both general and more specialized names information;
- the use of, and interest in, various nomenclature products and services--including those products and services presently offered by the province.

* Jeffrey R. Ball, Cartographic Toponymist, Ontario Ministry of Natural Resources.

An open question was also included with the questionnaire which allowed users to suggest other nomenclature products they would find useful; and a final

query was added to identify those organizations who currently maintain automated records incorporating geographic names.

The questionnaire was distributed to twenty-one major users of the provincial names record, identified from statistical data compiled during and since the feasibility study. In all, eight primary and thirteen secondary users received surveys.

Primary users who received a questionnaire included three toponymic agencies and five provincial departments engaged in either mapping, drafting, or surveying. Secondary users who also received questionnaires were: provincial departments involved in mapping--1; organizations responsible for managing Ontario's natural resources--11; provincial ministries which disseminate general government information--1.

Response to the survey was overwhelming. Seventeen or an unbelievable eighty-one percent, of the questionnaires were completed and returned for analysis. I take this opportunity to thank those who participated.

SURVEY FINDINGS

Current and projected Requirements for Information

Based on the results of the questionnaire, it would appear that major users of the provincial names record place a high priority on:

- recorded populated place and approved geographical feature names (see Figure 1);
- a graphic presentation of the toponymic extent of those named features and places contained in the provincial map record (see Figure 2).

The survey also provides considerable evidence to indicate that users generally obtain this information from maps, as Figure 3 attests.

In addition, the questionnaire furnished us with some useful data on projected needs for geographic nomenclature. The results suggest that organizations

who require geographic names information, may in the future, accept this kind of information in forms other than those currently used. These results were not conclusive, however, and this matter will be re-examined at a later time.

Additional Findings

Figures one through three disclose additional evidence on the current demands for geographic nomenclature. As we can see from the bar-graph in Figure 1, there does not seem to be a very strong demand for what might be classed as "typical" toponymic data at this time. You will note that alternate names, former names, local names, naming dates, generics, and origins all received a combination of fairly low "very useful", and high "do not use" responses. This result may be somewhat misleading, however. In point of fact, generics and decision dates both received high ratings from the primary users of the data base.

Demand for more specialized information is depicted in Figure 2. Notice that there is not a high demand for some of the more common, non-graphic methods of conveying information on the location, and extent of named features. You will notice from Figure 2 that coordinate data (i.e. a mathematical method of conveying feature location and extent), location descriptions, and drainage narratives (i.e. two textual methods of conveying this information), all scored rather low in comparison to the graphic method of feature portrayal.

A STRATEGY FOR IMPLEMENTATION

There are many different strategies which can be considered in building an automated geographic names data base. Based on the findings of the later user needs survey, the following strategy for automating the Ontario geographic name record has been recommended:

- that the province of Ontario automate the current geographic names record using the map portion of the record as its starting point: one way of doing this would be to enter into the automated record the minimum number of coordinates required to accurately convey the toponymic extent of named features and places as they appear on each 1:50 000 NTS map sheet; and/or as they appear on each 1:20 000 (and/or 1:10 000) OBM map sheet;
- if it is not feasible to automate all aspects of the graphic record, it should be determined which aspects of the graphic record most easily can be computerized: which will provide information on the location and extent of approved named features, and recorded populated places (approved for use); which will provide faster access to information presently held in the manual system; which will facilitate revision of the record; and which will provide the greatest degree of compatibility with the province's georeferencing data base;
- further, in addition to those graphic aspects to be computerized the initial automated data base should contain all approved geographical feature names and all recorded populated place names contained in the provincial record; and

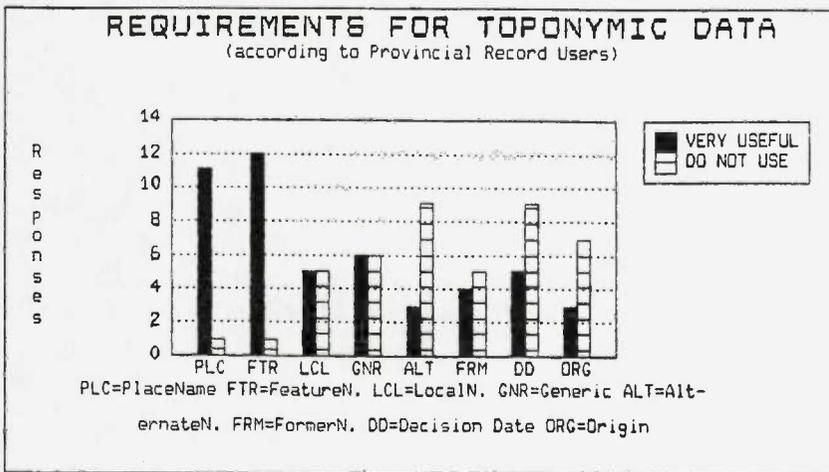


Figure 1 Graph showing the response to various kinds of toponymic data.

d) that the province consider including in the data base--generics, decision dates, and geographical township references for all the approved and recorded names referred to above.

RATIONALE FOR A "MAP-CENTERED" APPROACH TO AUTOMATION

A few of the more significant reasons for developing an automated geographic names data base from the map records of the current system are:

- a) major users of the record will be effectively served by a "map-centered" approach to automation because the present provincial geographic name map record contains the information these users most often require, including approved and recorded names, and the descriptions and locations of the features and places to which the names apply;
- b) the map records have been the primary user source for most of the information which is presently held in the card index, in the "Gazetteer of Ontario", and on Ontario Geographic Names Board decision lists. It would appear that so far the map records have been a good, and enduring record of geographical nomenclature;
- c) an automated graphic-based toponymic system will have the capability of generating not only graphic, but non-graphic products and information as well: such as an Ontario Directory of Geographic Names, and location descriptions, and drainage narratives;
- d) toponymic data bases which incorporate a reasonable amount of information on the graphic descriptions of features and places are likely to be more compatible with the powerful digital mapping data bases currently being developed, than are other systems which do not contain such information.

In summary, a toponymic data base built around the stability of a map record (which I suggest is the traditional user source of geographic names information)

would not only satisfy user needs now, but would be flexible enough to be expanded to satisfy additional needs in the future.

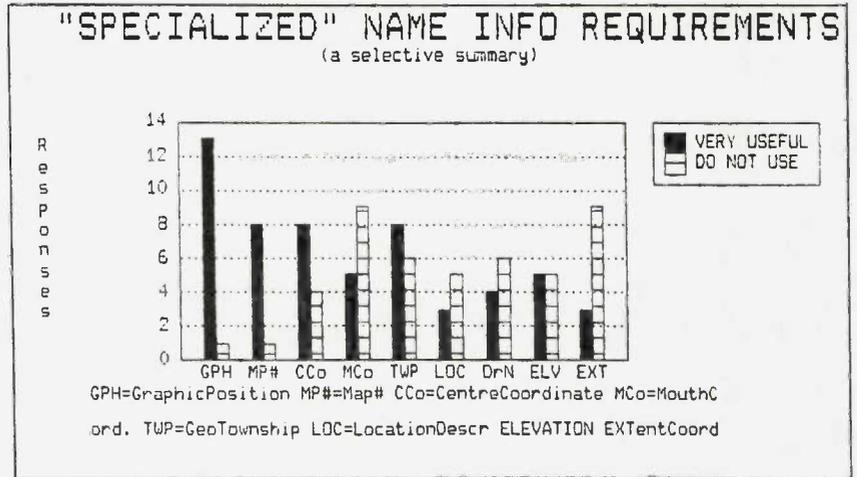


Figure 2 Graph showing the response to more specialized types of names information.

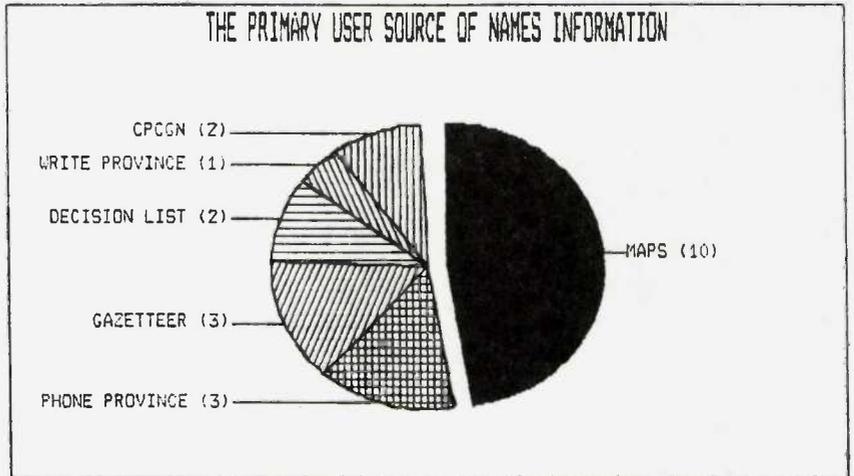


Figure 3 Graph displaying the results of preferred choice "balloting" for various user sources of information. Note that cumulative first-choice "ballots" are shown in brackets beside the product and service types identified.

APPENDIX A: QUESTIONNAIRE ON THE USE OF GEOGRAPHIC NOMENCLATURE

Please find attached:

- (1) A questionnaire dealing with the use of geographic nomenclature;
- (2) An example of an Ontario Geographic Names Board decision list; (see questionnaire - questions 1.1, 1.2, and 2.3)
- (3) An example of a "Graphic Index" showing approved Geographic Nomenclature of Ontario (see questionnaire - questions 1.1, 1.2, and 2.3)

Please complete the questionnaire and return it in the stamped, self-addressed envelope provided with this package.

QUESTIONNAIRE - USE OF GEOGRAPHIC NOMENCLATURE

Your Name :
 Organization :
 Department :
 Address :
 Telephone Number :

Please answer the following questions from an organizational rather than a personal point of view (ie. how you perceive your organization uses geographic names rather than how you personally use this information).

1.1 CURRENT USE OF PROVINCIAL NAMES RECORD

Please check one appropriate response for each question asked.

Question	Yes	No	Don't Know
(i) currently use Ontario Geographic Names Board decision lists?			
(ii) currently use graphic indexes?			
(iii) currently contact Nomenclature Section by phone for information?			
(iv) currently contact Nomenclature Section by letter for information?			

1.2 Please circle (below) those products and/or services you would like to use or continue to use.

OGNB Decision Lists

Graphic Indexes

Phone Nomenclature Section (MNR)

Letter to Nomenclature Section (MNR)

2.1 CURRENT USE OF OTHER SOURCES OF GEOGRAPHIC NAMES INFORMATION

Please check an appropriate response for each question asked. Note that blank spaces have been included if you wish to include additional information.

Where does your organization currently obtain geographic names information...?	Yes	No	Don't Know
(i) from maps?			
(ii) from Gazetteer?			
(iii) from the Federal Names Agency?			
(iv) from !			

Where does your organization currently obtain geographic names information...?	Yes	No	Don't Know
(v) from !			
(vi) from !			

2.2 Please specify - in order of preference - those products and/or services from questions 1 and 2 your organization finds most useful.

3.1 CURRENT AND FUTURE NEEDS FOR FUNDAMENTAL NAMES DATA

Please check one appropriate response for each type of data listed. In this question we are interested in determining your current needs for geographic names information. Question 3.2 will be devoted to determining your future needs for this kind of information.

Data Type Needed	Very Useful Now	Useful Now	Use Occasionally Now	Do Not Use Now
(i) Place Names				
(ii) Feature Names				
(iii) Alternate Names (ie. from another language)				
(iv) Former Names				
(v) Local Names				
(vi) Naming Date				
(vii) Generic (ie. where generic is not obvious eg. Red Lake (Place)).				
(viii) Origin (ie. name history)				

3.2 Please look at those data types you checked as "do not use now" or "use occasionally now". Of these types, please specify the one(s) you believe may be useful to your organization in the future. (If you wish, use roman numerals below rather than writing the data type in full).

4.1 CURRENT AND FUTURE SPECIALIZED NAMING NEEDS

Please check one appropriate response for each type of information listed. Use the back of this page (and next) to clarify where necessary or if you have any additional comments pertaining to the information. (Please include the roman numeral on the back of the page corresponding to the information type you are elaborating upon). Again, we are interested in determining your current needs for geographic names information. Question 4.2 will be devoted to determining your future needs for specialized names information.

Information Type Needed	Very Useful Now	Useful Now	Use Occasionally Now	Do Not Use Now
(i) Name and graphic portrayal of name application (ie. map showing names and toponymic extent of named features.)				
(ii) Map number on which a specific named feature appears [Please specify series].				
(iii) More than one map number reference (ie. if feature extends onto adjoining map sheets in the same series).				
(iv) Coordinates for centre of named features/places. [Please specify the system of coordinates you require (ie. geographical, UTM?) and the degree of accuracy (ie. nearest 0, ', ", 10,000m, 1,000m, or 100m)].				
(v) Coordinates for <u>mouth</u> of linear drainage features.				
(vi) Coordinates for <u>source</u> of linear drainage features.				
(vii) Coordinates to define <u>extent</u> of features/places.				
(viii) A <u>geographic</u> township reference for the centre of a feature/place.				
(ix) Multiple geographic township references (ie. if feature extends into more than one township).				
(x) A county or district or regional municipal or district municipal reference for the centre of a feature/place.				
(xi) Multiple county or district or regional municipal or district municipal references for features/places.				
(xii) Location <u>description</u> as given in Gazetteer (ie. Cronin Lake is NW of Trout Lake).				
(xiii) Location description with distances (ie. Cronin Lake is 1.6 km/1 mile NW of Trout Lake).				
(xiv) Drainage narrative (ie. Little Creek flows NW into Big Lake).				
(xv) Elevation of named features/places.				

4.2 Please look at those information types you checked as "do not use now" or "use occasionally now". Of those types please specify the one(s) you believe may be useful to your organization in the future. (Again, list roman numerals below rather than writing the information type in full).

5.1 USER INTEREST IN POTENTIAL GEOGRAPHIC NAME PRODUCTS AND SERVICES

In the question below we are attempting to determine your organization's interest in POTENTIAL geographic name products and services. Please understand that the products listed below are not currently available from the provincial names organization and will not necessarily be available in the future. The information obtained from this question however will be used in planning an effective geographical names information service.

5.2 Please check one appropriate response for each product listed.

<u>Potential Product</u>	Would Be Very Useful	Would Be Useful	Would Use Occasionally	Would Have No Use For
(i) Alphabetical lists of geographic names with corresponding centre/mouth coordinates (only) - sorted by the area of your choice. <u>Note:</u> List would include only those named features and places whose centre/mouth (ie. centre for lake/mouth for rivers) were located in the area chosen.				
(ii) Alphabetical lists of geographic names with multiple feature extent coordinates (if required) - sorted by the area* (approximate) of your choice. <u>Note:</u> Lists would essentially include <u>all</u> named features and places (if required) that occur in the area specified - including a few features and places which lie immediately outside the area chosen.				
(iii) Name lists similar to (i) and/or (ii) sorted by generic naming date or both.				
(iv) <u>Microfiche</u> lists similar to those preceding.				
(v) Map overlays displaying names, generics and feature extent coordinates. Area and scale of map to be selected by the user.				
(vi) Geographic names information in machine - readable form.				

* Area used would be defined in terms of whole Ontario Basic map sheets and therefore would be slightly larger than area chosen.

5.3 Please specify any one other nomenclature product or service your organization would find substantially useful.

6.1 Do you maintain an automated record containing geographical names data? Please circle the correct response.

Yes No Don't Know

THANK YOU FOR YOUR PARTICIPATION.

TOWARD A NEW SYSTEM

Commission de toponymie du Québec*

[Translation]

ABSTRACT. Mainly because of the prohibitive costs, the Commission de toponymie could not put the QNOMA system into operation. Instead, the Commission decided to improve the automated system which had been used for some twelve years. The improved system was named STRATO, "système de traitement de données toponymiques officielles" (automated system of official toponymic data). The main functions of this system are: correction, selection, statistics, official gazette, tables, fichier de toponymes populaires (file of unofficial locally used toponyms) and document management. In 1986 the Commission will have microcomputers in operation to meet its own management information requirements and to improve service to the public. The technological transfer to microprocessing will provide autonomy of operation, with almost immediate on-line data management.

BACKGROUND

At the first symposium on the automation of geographical names held in Ottawa in 1980, the Commission de toponymie presented a paper entitled SYSTÈME AUTOMATISÉ DE GESTION DE DONNÉES TOPONYMIQUES (computerized toponymic data management system).¹

The Commission's representative, Pierre Barabé, described in detail the features of the QNOMA system. As we made clear at the time, this was a project whose implementation could be quite different from what was anticipated - a prediction which proved only too true with time. In fact, although the Commission received approval in principle from the Quebec Treasury Board for the general structure of the QNOMA system, it was still to continue studying the mechanisms to be put in place to make possible the integration of the various existing toponymic systems, the resources required to integrate the data in the Commission's own data bank and, in particular, it was to conduct a cost-benefit analysis with a view to selecting software.

The months following the first symposium proved decisive in terms of the orientation that the Commission could expect to give its QNOMA project. It had to face the fact that it simply did not have the means or, more specifically, the financial resources to acquire the QNOMA system. It turned to outside consultants, who confirmed that the investment required to start up, develop, maintain and operate the QNOMA system was indeed prohibitive, particularly at a time of cutbacks.

TOPONYMIC NEEDS

The Commission de toponymie was, however, still faced with having to administer all the place names within the territory of Quebec, making them official, making them known to the public, and so forth. It had to exercise

its jurisdiction with regard to the criteria for selection of toponyms, the rules for writing the latter, geographic terminology, and so on. Indeed, far from decreasing, the commission's requirements in terms of technical support were growing apace. For example, on April 30, 1985, it had 122 133 toponymic records in its master file, each of which could generate a maximum of 23 data fields (Appendix 1), approximately a 30% improvement over the past five years. The master file, which is growing continuously, now handles a total of 2 578 599 data fields.

Thus, having had to postpone the QNOMA project, the Commission opted to upgrade the data processing system which, for some twelve years, it had used to produce the Gazette officielle du Québec and a number of lists of official toponyms arranged according to various parameters, including FEATURE, NAME, LOCATION, MAP number in the National Topographic System (Appendix 2). This second generation data processing system was dubbed STRATO, which stands for "système de traitement des données toponymiques officielles" (official toponymic data processing system). To improve performance and keep up with its growing toponymic data processing needs, the Commission established a new administrative unit called the SERVICE DE DÉVELOPPEMENT DES SYSTÈMES (Systems development service). To the day-to-day demands imposed by the ongoing administration of toponymy had been added the operation of a specialized file for a particular type of toponyms, TOPONYMES POPULAIRES (unofficial, locally-used toponyms) and a system to meet certain DOCUMENT MANAGEMENT (SGD) needs.

SYSTEMS DEVELOPMENT SERVICE

The primary activity of the Systems development service is the maintenance and operation of the existing data processing systems of the Commission de toponymie, which comprise the official toponymic data processing system (STRATO), the fichier des toponymes populaires (FTP) and the système de gestion documentaire (SGD) (document management system).

STRATO

Commonly called the toponyms file, this system is somewhat similar to the Répertoire toponymique offi-

* Presented by Jean Poirier, Assistant to the President, Commission de toponymie du Québec.

1 See CANOMA, Vol. 6, No. 1, July 1980.

ciel du Québec. Thus, it contains all the toponyms which appear in the *Gazette officielle du Québec* and, as of April 30, 1985, it contained 122 133 toponyms and odonyms (names of communication routes). While the primary purpose of this system is to produce the *Gazette officielle du Québec* each year, many other reports, including statistical reports and selective lists, are also produced.

All programs and procedures fall under the following six major functions:

- SUPPLEMENT

The purpose of the SUPPLEMENT function is to add records to the master file. This operation is two-fold: the content of the files is first entered at the terminal and stored up in a capture file, and the content of the latter is then added to the master file.

- CORRECTION

The CORRECTION function is used to make changes in the records of the master file. This operation is also two-fold: the corrections are first captured at the terminal in a transaction file and these transactions are then processed by the updating program, which makes the corrections in the master file.

- SELECTION

The SELECTION function extracts records from the master file to produce lists of toponyms or extracted records or simply to count them. The selection criteria are limited only by the number of fields in the file.

- STATISTICS

The STATISTICS function produces statistics on the content of the master file or any other file extracted from the master file. The statistics produced are simply cumulative totals based on various categories of data.

- OFFICIAL GAZETTE

The OFFICIAL GAZETTE function serves to produce a file that can be used by a typesetting firm. This file contains the information which is to be published each year in the *Gazette officielle du Québec* (namely, toponyms which have been made official since the last publication and any changes made to existing official toponyms).

- TABLES

The TABLES function combines all operations affecting coding tables: creation, updating, and listing. The system comprises six tables: township, census division, feature, geographic code, county regional municipality, and administrative region.

FTP (Fichier des toponymes populaires)

The purpose of this file is to respond to requests for toponymic information at the level of the territorial divisions of Quebec.

For the period relating to the 1984-1985 fiscal year, more than eighty reports were distributed to the users of this service offered by the Commission de toponymie.

SGD (Système de gestion documentaire)

The document management system was established in 1983-1984 to save and update files containing descriptive data on documents used by the various administrative units of the Commission.

For the period relating to the 1984-1985 financial year, more than 1 300 transactions were carried out.

CHANGES TO THE SYSTEM

Certain changes have been made to the existing data processing system to enable the Commission to complete the official toponymic data already available for Quebec.

One of the main improvements made to the system was the creation and updating of the tables of geographic codes (municipalities), county regional municipalities, and administrative regions.

A few adjustments were made to existing functions in the master file in order to optimize the present status of this file. These adjustments include the addition of a second pair of coordinates for greater precision, the addition of the feature category, which means that it is now possible to distinguish between the record of an odonym and that of a toponym, and certain minor modifications to the census division and the townships.

MICROPROCESSING

The Commission de toponymie is in the process of researching and studying its data processing options in order to find a comprehensive technological and organizational approach to meeting its needs in a context of data management and enhanced service to the public.

The adoption of a microprocessing approach will involve firstly the transfer of all data contained in the existing system (STRATO) to microcomputer, with all such a transfer entails in terms of actual facilities, software, the training of personnel and the establishment of a security system.

As the changeover is initiated, the parallel use of both systems will enable the Commission to continue its normal activities (capture and updating on the old system) while retaining a margin of safety during the period of adaptation to the new system.

Secondly, the two data files will then be compared and the necessary adjustments made, with the old system finally abandoned as soon as the new one is broken in.

This autonomy of operation combined with almost immediate on-line data management, the continuity of project development, the possibility of using software adapted to the needs of the Commission, and the reduction in short-term and long-term operating costs are only a few of the technological and organizational advantages to the Commission de toponymie of the technological change-over to microprocessing.

We hope that this rational approach will be part of the day-to-day experience of the Commission de toponymie by the end of the coming year.

APPENDIX 1

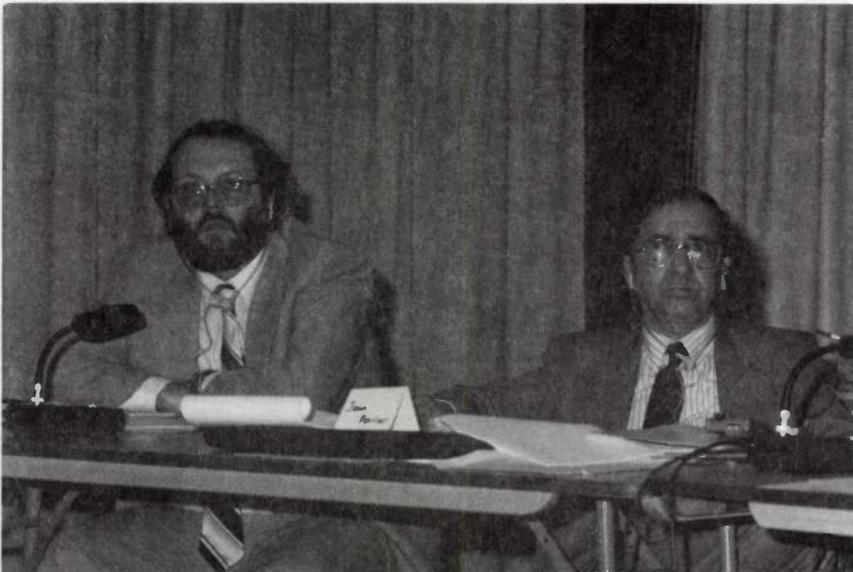
TYPES OF TRANSACTIONS AVAILABLE IN THE SYSTEM

<u>Transaction</u>	<u>Before 1985</u>	<u>1985</u>	<u>Transaction</u>	<u>Before 1985</u>	<u>1985</u>
Sequential number	1	1	Category of feature (odonym/toponym)	0	2
Co-ordinates	1 pair	2 pairs	Census division	1	4
Linguistic coding	1	1	Geographic code	0	4
Semantic code	1	1	Township	3	4
Publication date	1	1	Toponym	1	1
Feature code	1	1	Map sheet	1	1
			Cross-reference	1	1

APPENDIX 2

ADDITION OF NEW TOPONYMS TO THE FILE

- Data to be captured (minimum):
- toponym
 - map sheet number
 - feature code
 - category of feature (type: odonym/toponym)
 - co-ordinates (one pair)
 - geographic code (minimum 1; maximum 4)
 - census division (minimum 1; maximum 4)
 - one cross-reference
- Supplementary data:
- township
 - date of publication
 - co-ordinates (2nd pair)
 - linguistic coding
 - semantic coding



SYMPOSIUM PARTICIPANTS:

On the right: Jean Poirier (Commission de toponymie du Québec)

On the left: Clément Nolette (Université Laval)

AUTOMATED LOCATION OF TOPONYMS IN THEIR RESPECTIVE MUNICIPALITIES

Clément Nolette*

[Translation]

ABSTRACT. This text describes the approach taken to the automated location of Quebec toponyms in their respective municipalities on the basis of the geographic co-ordinates of the toponyms and a numerical file of the municipal boundaries.

At the request of the Commission de toponymie du Québec, we have developed a program for the automated addition of the municipal codes to the toponyms in their data bank. The geographic co-ordinates for the locations of the toponyms were already part of this data bank and we knew that the digitized base map of the municipal boundaries of Quebec had recently become available from the Department of Energy and Resources. Thus, all the prerequisites for this project were in place and all that remained to be done was to establish the link between the two files. In the interest of efficiency, we divided the work into two phases, one to involve the preparation of modified files and the other for the processing of data. In the second phase, to ensure accuracy, we introduced a mechanism whereby the results obtained could be classified in three different files according to their reliability. A brief description of the algorithms used follows the description of the system.

In the first phase, we started with a version of the toponyms file which contained no toponyms or hydronyms, because of the linear nature of what they represent, and no toponyms for which the municipal codes were already present. From this file we extracted only the information we needed in order to limit the size of the file processed. Thus, for each toponym we retained its reference number, its co-ordinates, the code for its census division, the number of its map sheet and the first fifteen characters in its name. In order to facilitate the mathematical treatment to follow, however, we modified the form of the co-ordinates and used degrees and fractions in place of degrees, minutes and seconds to represent latitudes and longitudes. The number of the map sheet and the first characters of the name were retained in order to facilitate the task of the person who would verify difficult cases which the program would have been unable to classify with certainty.

The file produced by this extraction process is then sorted by order of increasing census division number. All the toponyms are thus grouped in consecutive sets corresponding to each of the divisions (see Figure 1). The same sorting process will then be used on the municipalities file, thus making it possible to process the toponyms on a division-by-division basis. The purpose of all of this is to reduce the amount of research time by limiting the number of municipalities which the program must search through for a given toponym.

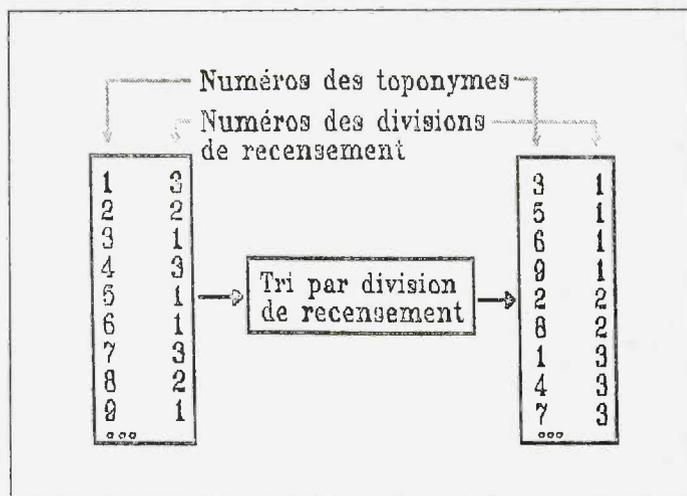


Figure 1

The municipalities file is a F.I.L.A.M. (Fichier informatisé des limites administratives municipales) "computerized file of municipal administrative boundaries" file, supplied by the Quebec Department of Energy and Resources. This file contains all the municipal boundaries in the province and each municipality is represented by a polygon which is assigned a geographic code. This is precisely the code which we wish to add to the files of the toponyms located in each municipality, in each polygon. For the municipalities file we have also transformed the geographic co-ordinates, latitude and longitude, into degrees followed by decimal values rather than degrees, minutes and seconds. Spherical co-ordinates are then used directly and treated like cartesian co-ordinates since the exclusion or inclusion of a point in a polygon would not have been changed by the algorithm for the projection that could have been applied to the co-ordinates for the toponyms and for the municipalities.

Before proceeding to sort the municipalities file by census division, we carried out certain calculations to obtain, for each municipality, the boundaries and the area of the square circumscribing its polygon. This new data was added to the municipalities file and during the division-by-division sort, a second sort was carried out within each division by order of increasing surface area of the circumscribing squares. The need for this second sort can be explained by the manner in which the search is carried out to locate a toponym in

* Clément Nolette, graduate student, Department of Geography, Université Laval.

one of the municipalities of its census division. This search is carried out sequentially, and in order to ensure that toponyms located in a municipality which overlaps another larger municipality are not located in the latter by mistake, we must be certain that toponyms are checked for inclusion in municipalities of small surface area before those of larger surface area (see Figure 2).



Figure 2

Having completed the preparation phase for the toponym file and the municipalities file, we move on to the processing per se (see Figure 3). This produces three files, all of which contain the data retained during the extraction carried out in the first phase as well as the appropriate municipality code if the search has proceeded smoothly. The first file contains the accurately located toponyms, or those which have been located with certainty within a municipality. The second contains the less accurately located toponyms, or those which have been located within a certain critical distance of the boundaries of the municipalities found for these toponyms; this distance is parametric and may be modified as required. The toponyms placed in this file should normally be verified manually. Finally, the third file contains the toponyms which could not be located in any municipality, namely those belonging to the T.N.O. (Territoires non organisés) "non-organized territories". Certain statistical data is also supplied by the program, including the number of accurately and less accurately

located toponyms per municipality as well as the number of toponyms per census division which belong to the T.N.O.

The following is a brief description of how the program functions. It begins by loading into memory all the data on the municipalities of a given census division: geographic codes, circumscribing squares, surface areas and co-ordinates. It then reads the toponyms one by one, processing them as it does so. When it encounters a toponym for which the census division code differs from the municipality code, it then reads the data for the municipalities in the next division before resuming the processing of the toponyms. This process continues until the end of the municipalities file and the end of the toponyms file are reached.

A toponym is processed by verifying whether it is located in one of the municipalities in its census division. Each municipality is checked until the one in which it is located is found. If no municipality is found, the toponym belongs to a T.N.O. An initial test reveals quickly whether or not a point is located within a given municipality: it is simply a matter of checking to see whether it is located inside or outside its circumscribing square (see Figure 4). If it is located outside the square, it is clearly not within this municipality and the next municipality is checked.

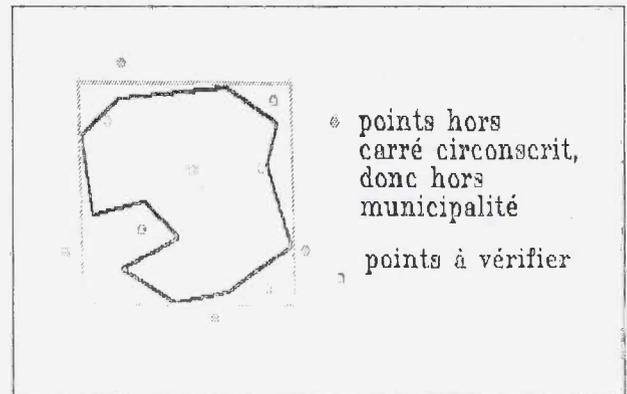


Figure 4

On the other hand, if it is within the square, a more complex test to determine whether or not a point is located within a polygon, described in the following paragraph, is carried out. If the point is not within the polygon, the next municipality is checked. Alternatively, the distances from the point to the various segments making up the boundary of the municipality and to the vertices formed by the intersection of such segments in the case of angles in excess of 180 degrees are calculated. If a distance which is less than the critical distance given as a parameter is encountered, the toponym will be classified as less accurately located; otherwise it will be considered to be accurately located (see Figure 5).

The algorithm used to verify whether a point is located inside a polygon is based on the principle that a straight line crosses either none or an even number of the sides of a polygon located in the same plane, unless it goes through a vertex. Thus, we draw a vertical line through the point whose location is to be verified and see first whether it also passes through one of the vertices of the polygon. If so, a slight shift of the

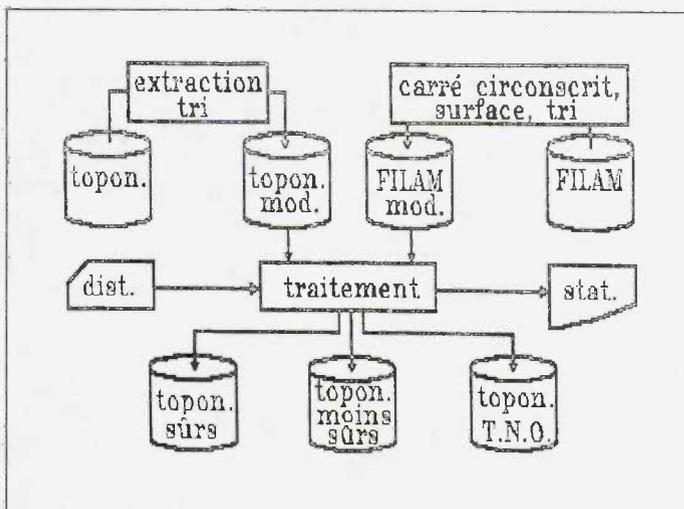


Figure 3

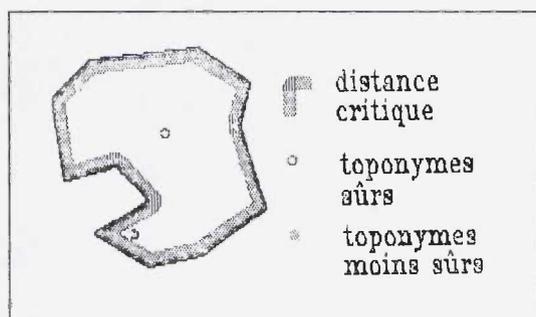


Figure 5

abscissa of the point in question allows us to continue the test, after repeating the above verification. We then calculate the points of intersection of this vertical straight line with the various segments of the polygon that it crosses. Finally, we count the number of points of intersection located above or below the point. If this number is odd, the point is located inside the polygon. If the number is even, the point is outside the polygon (see Figure 6).

At the time of writing, the system had been tested with twenty-three municipalities in two census

divisions of the Charlevoix region and all the toponyms of the region. All the toponyms classified as accurately located are, in fact, accurate, while in the case of over eighty per cent of the less accurately located toponyms, the general locality was correctly found. It is thus a matter of calibrating the critical distance to ensure that the latter toponyms are accurately located and at the same time that no doubtful cases are included. A few toponyms classified as belonging to the T.N.O. have no reason to be there and we are currently searching for a solution to this problem.

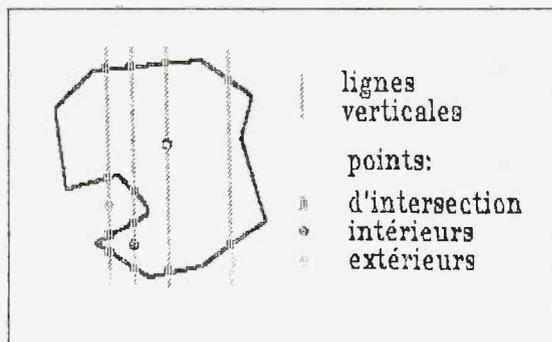


Figure 6

GEOGRAPHIC NAMES INFORMATION SYSTEM (GNIS):

PHILOSOPHY AND FUNCTION

Roger L. Payne*

ABSTRACT. The Geographic Names Information System (GNIS) is an automated names system that is the result of long-term efforts aimed at the collection and publication of geographic names. It is a multipurpose system designed to meet national needs by providing services to a wide range of users at all levels of government and the private sector. The system has been designed by names experts to serve as the national names depository of the United States, to assist the National Mapping Program of the U.S. Geological Survey (USGS), and to provide a basis for toponymic research and application. The system will be maintained on a continuous basis, and it offers a wide range of products.

INTRODUCTION

The collection and publication of geographic entity names in the form of gazetteers and geographic

indexes has been an expedient means to identify places, features, and areas of the United States since the 18th century. Since geographic names are needed as references and for identifying and relating locations, the growing numbers of names produced in connection with the expansion of settlement during the 18th and 19th centuries soon precluded any degree of published completeness even on a regional or local scale. A large number of State and national gazetteers were published in the 19th century,

* Roger L. Payne, Chief, Geographic Names Information Section, National Mapping Division, United States Geological Survey, Reston, Virginia.

but these primarily made reference to populated places and larger physical features. The matter of completeness was generally recognized as difficult or impossible to attain. In 1892 the U.S. Geological Survey began cataloging geographical names as part of the national mapping effort, and gazetteers of 12 States and 3 territories were published between 1894 and 1906. While accuracy and completeness were always goals, these gazetteers only reflected the names found on the topographic maps of the Geological Survey, which were basically at scales of 1:125,000 (a 30-minute area) and 1:63,360 (a 15-minute area). Many local names for minor features were omitted. Furthermore, only a small percentage of the country was mapped by the beginning of the 20th century. Thus, inadequate map coverage led to the discontinuation of the program.

FORMULATION OF A DIGITAL DEPOSITORY

The philosophy of the development of an automated names system is embodied in past attempts, current needs, and projected requirements. Through the first half of the 20th century the large-scale topographic map coverage of the United States increased considerably, but not to the extent that allowed a systematic collection of geographical names for gazetteer purposes. There were some local and regional activities, but not organized on a national scale, and often the work was carried out by untrained people. An effective gazetteer should list more than a name and its broad, relative location, such as State and county. Specific named feature positions identified by a grid system, such as latitude and longitude determined to a precision level of seconds, were needed. Similarly, a reference to the map sheet on which the feature is located would assist in quickly identifying locations. Additional information in the form of variant or secondary name usage would assist in identifying historical references and applications, as well as providing insight into the toponymic evolution of the feature. Precise location was not considered in early gazetteer efforts because it was thought unnecessary for the products being produced. To be complete, a listing of names must rely on a number and variety of sources from all levels of government and especially historical materials.

As early as 1953, Lewis Heck of the former U.S. Coast and Geodetic Survey called for a National Gazetteer that would be as complete as possible, relying on material from a wide variety of sources and that should include "sufficient details to provide a rapid location of any name and preferably should list variants in cases of name conflicts."¹

From the mid-1950's through 1970, there were many attempts both in and out of government in the United States to establish comprehensive names listings (some digital), but their usefulness was limited because of the level of precision and their lack of completeness. Many of those attempts were made to develop digital files capable of some rudimentary data handling, but the technology was not yet available for complete development and processing of geographical names files.

A single, unbiased, relatively complete depository of geographic names, including basic or essential

information about each name and named place, was needed by a variety of users. Because of the nature and size of the task, most users looked to the Federal government for help in developing and maintaining a master computer-based depository of the Nation's 3 to 5 million domestic names.

The U.S. Geological Survey agreed in 1979 to undertake such a project as part of the National Mapping Program and to assist the national standardization program of the U.S. Board on Geographic Names. The USGS had been experimenting with automated names processing since the early 1970's, and a prototype data base was developed. Improved technology in data retrieval and ancillary data storage allowed implementation of the National Geographic Names Data Base (NGNDB). Further improvements and enhancements led to the development and implementation of a totally automated names system called the Geographic Names Information System (GNIS). GNIS is a system of data bases, data retrieval software, and procedures specifically designed to function as an official geographical names depository and as a tool for toponymic research.

The computer-based system is designed to meet a broad spectrum of information and program needs, including:

- . cartographic support
- . standard reference
- . geographic base for special files
- . national standardization
- . toponymic and other research
- . digital application
- . geographic inventory and control
- . gazetteer production
- . special publications

Information in the system may be retrieved, arranged, analyzed, and manipulated for general and specific purposes. The paramount and largest data base is the National Geographic Names Data Base, which contains a separate file for each State, territory, and outlying area in the United States, as well as various topical files. The other data bases in GNIS are currently:

- . National Topographic Map Names
- . Reference/Bibliography
- . Board on Geographic Names

BASIC INFORMATION

The effectiveness of GNIS and the National Geographic Names Data Base depends upon both the completeness and integrity of the data in the system. Research and experience led to the identification of certain basic or minimum data categories or elements needed for each name or record in the data base. These include the:

- . written form of the full official, legal, or standard name
- . designation of the kind of geographical feature
- . location by State and county (name and code)
- . location by topographic map (name and code)
- . identification and extent of the named entity by geographical coordinates
- . official status of the name and its application
- . variant or secondary references to the name
- . bibliographic reference to the source from which the name was obtained

In addition to the required elements, other

1 Heck, L. (1953): "The problem of a national gazetteer", Names, vol. 1, no. 4, December, p. 233-238.

data elements are being compiled or are being considered for entry into the data base. Such data include information relating to elevation, population, textual description, name origin and history, Federal land survey reference, and reference to ownership of administrative areas as well as codes referring to census, postal, and fluvial areas. Additional categories will be added as needed.

COMPILING THE DATA BASE

Defining necessary data elements and building the data base require organizing the work into several steps, tasks, and phases. The major task is the collection of names and associated information to go into the data base. Written sources of name information are extensive and varied, and include both current and historical maps and charts as well as textual materials. Only published name information is considered for building the data base.

There is no question that the largest and most uniformly distributed coverage of current geographic names of the United States and its territories is found on the topographic maps published by USGS. A survey made in 1970 by USGS indicated that the large- and intermediate-scale topographic maps contained from 50 to 70 percent of all names in current usage.

This situation provided a logical division of the work into two distinct phases:

- Phase I, the collection of names and associated information from the USGS standard topographic map series, and
- Phase II, the collection of such information from other sources.

Phase I

The Geological Survey began preliminary work on Phase I in 1976, when published large-scale topographic maps covered more than 70 percent of this country's land surface. In a pilot project, names published on the maps covering the States of Kansas and Colorado were recorded in machine-readable form. In the fall of 1978, after a favorable evaluation of the pilot project, a contract was awarded to record corresponding information for the remainder of the States and territories. The contract was completed in January 1982. All work was monitored closely for completeness, accuracy, and adherence to established procedures by the Branch of Geographic Names in the Geological Survey's National Mapping Division. The error rate was about 2 percent.

Phase I activity for building the data base involved the:

- systematic encoding of most names and attendant information from the topographic maps on a State-by-State basis, and
- correction and updating of these data by comparison with the official BGN files.

Phase II

To insure a data base as complete as possible, the names of features not recorded on topographic maps

and from a variety of sources, including historical material, need to be added. The systematic collection of names from other sources is termed Phase II.

The main thrust of Phase II work for each State project consists of a systematic search of selected documents to create new name records to add to those created during Phase I and to collect variant names for all name records. This work must be completed before a State or territory volume of the National Gazetteer of the United States of America is published.

Phase I compilation did not include all named entities shown on the topographic maps. For example, certain kinds of named features that were large or scattered could be handled more efficiently during later compilation, or the information was already compiled and was available on special lists or in digital form. Phase II compilation requires coding those name records omitted from Phase I compilation, and researching and recording names and associated information not found on the topographic maps.

As mentioned earlier, topographic maps do not show all geographic names that are in current use. Many names in local and (or) published usage were unknown to the compilers when the maps were made. Names applied to minor features often were not shown for lack of map symbols or to avoid map clutter. The maps also do not normally show historical and obsolete names, variant applications, and other names once used for currently named places, features, or areas. Complete instructions regarding the collection, documentation, and encoding of geographical names for Phase II compilation were published in The National Geographic Names Data Base: Phase II Instructions.²

SYSTEM DEVELOPMENT AND IMPLEMENTATION

One of the most important aspects of an automated names system is to be able to store and retrieve information with no preconceptions about the data and no introduced bias. Functionally, a data retrieval system for use with geographic names has to have this one quality above all others. Other desirable qualities are rapid retrieval time, efficient means of partitioning data, flexible output and report formatting, as well as low cost. An important requirement for a data retrieval system is that it should be "user friendly" or easy to learn for a wide variety of users. Additional requirements include ease of data and data category addition and the ability to append other software for specialized and highly sophisticated data manipulation.

Development of specific automated names retrieval software by GNIS staff would have required considerable time, effort, and funds. The USGS already had data retrieval software that incorporated virtually all the services identified for the implementation of an automated names system. The result was to adopt GIPSY³ as the

2 Orth, D.J. and Payne, R.L. (1984): The National Geographic Names Data Base: Phase II instructions, U.S. Geological Survey Open File Report 84-036.

3 GIPSY[®] was developed by Information Systems Program Department of the University of Oklahoma in Norman, Oklahoma. GNIS currently uses version 1.0 but is testing version 2.0 for future implementation.

data retrieval software. GIPSY allows direct communication between the user and the data base in clear English form which minimizes ambiguities. One may use one's expertise and training for data manipulation, which permits reflective or introspective research on the data without having to acquire more than a rudimentary understanding of information systems. The various files in an automated names system should be accessible simultaneously from multiple locations, which include the information outlets of Federal and State governments and eventually personal computers. GIPSY utilities allow virtually unrestricted addition of data and data categories. Additionally, there is no restriction on length of data nor variability of length, which is critical for an automated names system.

OPERATING ENVIRONMENT

GNIS operates in a mainframe environment, which means that it is multiprocessing and operates with virtual storage (MVS). A directory system is used for identifying valid entries within a record, and dynamic addressing of a temporary file is employed in the retrieval system. An index or inverted file may be created in GNIS if necessary, but repetitive searching in an automated names system is usually not an application. Access to GNIS is available interactively at all times by telephone connection. Users find a "user friendly" menu-driven system offering interactive (real-time) retrieval as well as batch retrieval, and in many places, a tape copying operation may be performed. Many users may begin utilizing the system after reading the GNIS Users Guide⁴ and following the step-by-step examples. A short training course may be provided for those interested in data structure and advanced retrieval techniques. For those without retrieval capability, the USGS provides user services for a nominal fee, which includes data retrieval, books, microfiche, and magnetic tapes. Wider access to GNIS will be available through the USGS Earth Sciences Information Network (ESIN). ESIN is a collection of software designed to allow users direct access to a variety of earth sciences data bases that reside on different mainframe computers. Currently ESIN is available only at

USGS information offices, but may be available to other users in the future.

MAINTENANCE

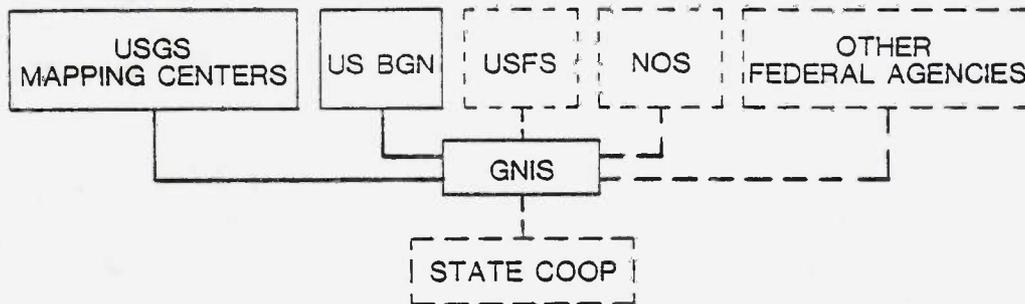
Part of the overall function of information systems is the structure and schedule of maintenance. The maintenance structure of GNIS is not completely operational, but the foundations of maintenance have been established and are partly operational (see Figure 1). GNIS is maintained by the National Mapping Division of the U.S. Geological Survey. Each regional mapping center compiles and formats new names data and minor corrections, which are electronically transmitted to the GNIS staff for a final check before the data are entered. The U.S. Board on Geographic Names transmits data directly to the GNIS staff for entry into the National Geographic Names Data Base. A series of checks and balances assures integrity and security so that all users can retrieve and use data with confidence. Maintenance is continual, and flexibility is built into the system for any needed adjustment. All users have retrieval capability, but only the GNIS staff has the capability to alter, add, or remove information.

While the maintenance program is currently operational in the regional mapping centers of the National Mapping Division and the U.S. Board on Geographic Names, the possibility of extending this capability to other Federal agencies and possibly some State agencies is being considered.

SUMMARY

The Geographic Names Information System is the result of a long-term effort by diverse groups to establish a single unbiased means of generating complete listings of geographical names. The results and research efforts of a number of experts in the field of toponymy have been applied to the development and implementation of the automated names system. The system is designed to serve the needs of a wide variety of users in all levels of government and the private sector who are engaged in research and application problems. The system also provides assistance to a national effort of standardization and mapping support. A variety of services and products are available and the system is maintained continuously.

4 Payne, R.L. (1984): Geographic Names Information System Users Guide, U.S. Geological Survey Open-File Report 84-551.



————— INDICATES AN OPERATIONAL PROCEDURE
 - - - - - INDICATES A PROCEDURE IN DEVELOPMENT

USGS = US Geological Survey
 BGN = Board on Geographic Names
 USFS = U.S. Forest Service
 NOS = National Ocean Service

Figure 1 The current maintenance structure of the Geographic Names Information System (GNIS)

TERMIUM III

John Carey*

The Translation Bureau first offered a computerized terminology bank to its translators and clients over nine years ago. The original TERMIUM Bank was developed by the University of Montreal in 1970. The Translation Bureau took it over a few years later, tailored it to its own needs and called it TERMIUM II.

It soon became evident that, because of the limited technical possibilities of this second generation system and the expanding network of TERMIUM users both in Canada and abroad, a new system would be required. Consequently, from 1982 to 1985, the Department of the Secretary of State developed its own third generation system (TERMIUM III) using the BASIS software package in conjunction with Telidon technology which permits the use of menus and colour graphics. A computer was also purchased in order to ensure that the Bank's needs would be served and to facilitate access to information by users.

data base contains the following four data collections:

- 1) Terminology Data Collection
- 2) Multilingual Data Collection
- 3) Translation Problems Data Collection
- 4) Proper Names Data Collection

The Proper Names Data Collection is certainly the one that will interest your association the most. It contains records whose entries are proper names. The most common examples are names of organizations, committees, programs and places, trade names and official titles of acts, regulations and documents. At the present time, the Bank contains approximately 6,100 records in the geographical names field (including toponyms). TerMIUM III is a menu-driven system providing menus in English or French depending on the language of the user. The following is an example of a record (see Figure 1).

```

                                ** Appellations - **
DOMAINE(S)                      : *Toponymes (Canada)
EN                               FR
Kicking Horse Pass*a           col Kicking Horse*b

OBS*Mountain pass, Canadian Rockies, on boundary bet. SE British Columbia and Banff National Park in Alberta.*a
OBS*col des Montagnes Rocheuses, à la frontière de l'Alberta et de la Colombie britannique; 1 628 m.*b

SOURCES
-----
a*WEGEO*1969***556
b*3DFX*1985
```

Figure 1

Following the conversion of data to the new system, the number of records accessible in the Bank totalled approximately 750,000 representing some 1.8 million terms.

The TERMIUM Bank now consists of a linguistic data base and a documentation data base. The linguistic

The record contains various blocks of information, among which are found basic elements such as the entry terms in each language, textual support, sources, subject fields, author and date. The data collection to which the record belongs is also indicated as well as any noteworthy comments about any of its terms.

Terminals providing access to TERMIUM serve not only the 1,200 Bureau translators, but also the entire federal public service, several provincial governments, universities in Moncton, Montréal, Toronto, Ottawa, Kingston, Quebec and Sudbury, various research institutes, translation companies and foreign organizations such as the Commissariat général de la langue française in

* John Carey, Officer, Client Services, Network Management, Terminology Directorate, Translation Bureau, Secretary of State.

Paris, the United Nations headquarters in New York, the International Monetary Fund in Washington and the International Maritime Organization in London.

Managed as a cooperative, the Bank is a common tool for all its users, who are not merely clients buying a service, but also suppliers contributing to the de-

velopment of a common data base by providing new records or upgrading existing data.

Like language itself and the terminology people use, the Bank is constantly evolving. It is a cooperative venture, and its progress will continue to depend on the cooperative efforts of all those who contribute to its operation.

CANADIAN HYDROGRAPHIC SERVICE REPORT ON THE AUTOMATION OF UNDERSEA FEATURE NAMES

Mr. Colin Bromfield* reported briefly on the automated system used by Canadian Hydrographic Service (CHS) to produce the "Gazetteer of Undersea Feature Names" (1983). In 1982, card records for some 3500 undersea feature names were automated on a PDP 11, through a programme written for DATATRIEVE software package. Later the data were upgraded onto a VAX 780. DATATRIEVE is not a data base management system, but a simple enquiry and reply system, biased toward low-end querying and reporting. The menu-driven system is user-friendly and, for this small-sized data base, asks questions to assist an inexperienced operator perform the main functions,

* Colin Bromfield, Analyst/Programmer, Canadian Hydrographic Service.

such as creating, editing or deleting records. DATATRIEVE is also being used to undertake much of the verification of data input. For French characters an 8-bit representation in ASCII format allows representation of the accented letters. The CHS names system was developed based on the experience of the NTDB; the basic design of a names record was modified and adapted to the needs of the undersea feature names data. Successfully the first edition of the gazetteer was generated in 1983. At the present time, additional material is being added to enlarge the historical and name origin attributes of each record. Mr. Bromfield indicated that the CHS data base had the flexibility to be plugged into other systems or reformatted for adaptation to a larger or better system in the future.

STATISTICS CANADA: ACTIVITIES USING GEOGRAPHICAL NAMES

Representatives of Statistics Canada made brief representations describing toponymy problems and issues encountered in their operations.

Mr. Henry A. Puderer* began with a general overview of the services provided by the Geography Division

* Henry A. Puderer, Chief, Spatial Delineation and Analysis Section, Geography Division, Statistics Canada.

in support of the statistical data collection, compilation and dissemination activities of Statistics Canada. These services or functions are identified in Figure 1 and the geographic framework supported for the 1986 Census of Canada is presented in Figure 2.

Mr. Puderer shifted his focus to toponymic issues affecting their operations, specifically (a) the representation of geographic names in upper and lower case, with accents, on files, maps and in publications of Statistics Canada; and (b) the differences in the classification systems of place names used by the Geographic Division and EMR.

- Geography Divisions 6 Principal Functions**
1. Geographical Concepts and Delineation of Areas.
 2. Census Mapping for Collection.
 3. Census Geographic Master File. (CGMF)
 4. Geographically Referenced Data Storage and Retrieval System. (GRDSR)
 - The Geocoding System.
 5. Tabular and Cartographic Products.
 6. Population Counts.

Figure 1

In the first instance Mr. Puderer urged that the development of standards be pursued for the machine representation of accents and protocols that would maximize the ease of toponymic information exchange.

In the second instance, Mr. Puderer explained that the differences in the way place names are classified by EMR and Statistics Canada currently confound Statistics Canada's efforts to match their place name files with the NTDB of EMR. While this matching is important to Statistics Canada so as to ensure that their publications present official place names, as approved by the CPCGN, it is also essential to EMR for data from Statistics Canada to be easily accessed, as for example, in preparing the Canada Gazetteer Atlas and the National Atlas of Canada. Statistics Canada will be pursuing this issue in cooperation with EMR.

Mr. Joel Yan** directed his comments to the

** Joel Yan, Chief, Methodology, Geocartographics Sub-division, Statistics Canada.

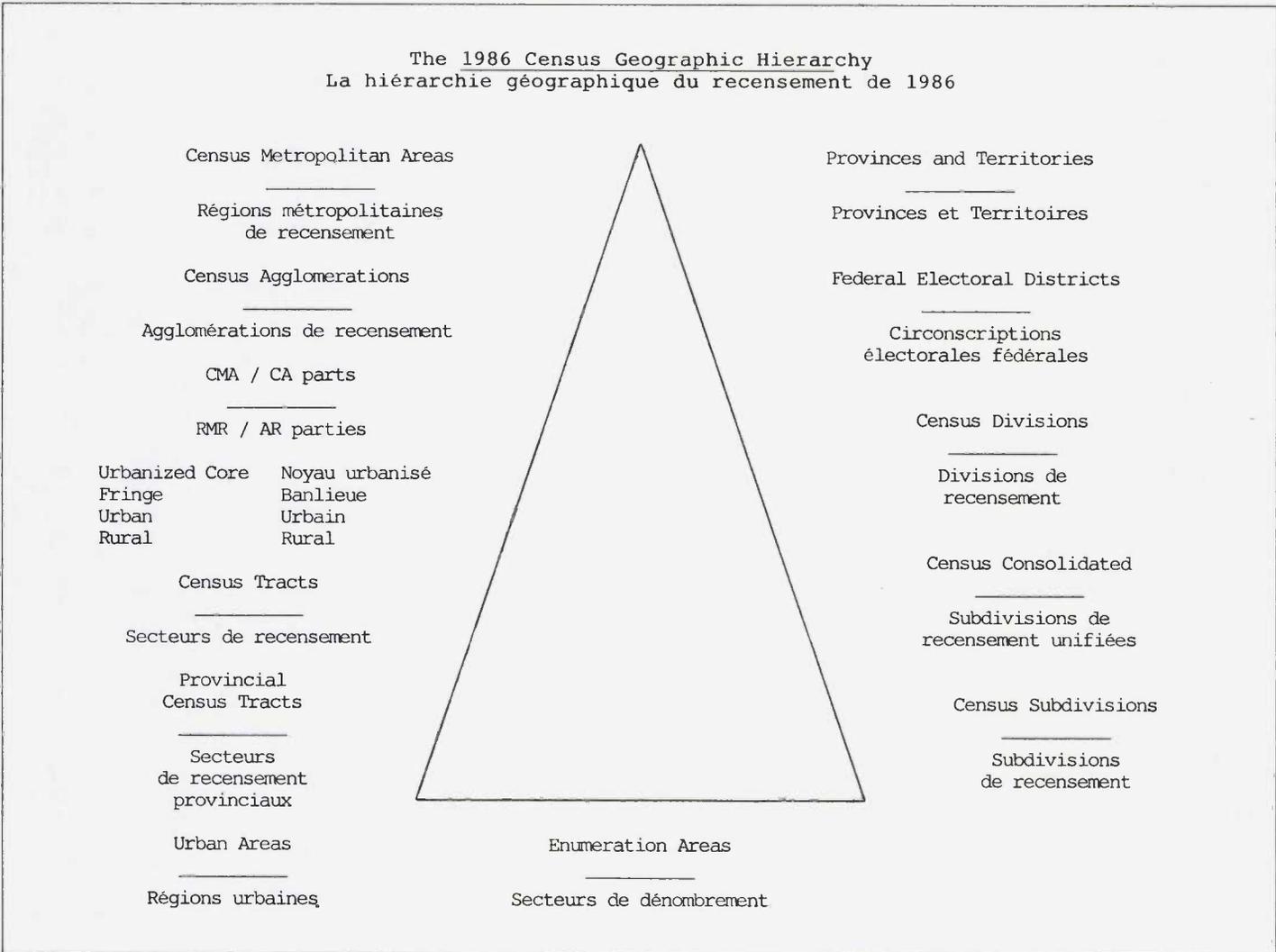


Figure 2

issues of Statistics Canada street name information and the automated placement of names on maps.

In order to provide a small area data retrieval capability in large urban centres, Statistics Canada maintains digital street network files, known as Area Master Files (AMFs). Current coverage of approximately 200 municipalities accounts for about 56% of the Canadian population. One component of the files is names for all features, e.g. streets, highways, railroad lines, hydrography and major boundaries. The names are compiled from municipal documents. These files are also used in the production of automated census data collection maps (Figure 3) as well as automated single line maps for private

and public dispatching applications (Figure 4).

In presenting these names several issues have arisen:

- (a) representation of special characters, e.g. accents;
- (b) formatting of the names; and
- (c) conventions and verification of street names.

The automated text placement for urban street maps is based on a computation of the text string length and a search for a street segment or pair of adjacent



Figure 3 Census Tract (CT) Base Map

After the AMF is disaggregated by CT, double line streets and block numbers are generated, and automated text and address placement is done for each CT file. The resulting plot is reproduced and used by Census representatives (enumerators) to ensure orderly and complete coverage of the delivery and pick up of questionnaires for the 1986 Census. Approximately 1,100 of some 3,800 census tract maps for 1986 are being produced by computer-assisted means.

The system was developed by Geocartographics Subdivision based on geographic inputs (Area Master File and Census Tract Boundaries) maintained by Geography Division.



Figure 4 Ambulance dispatch map
(Part of Model MH47 21/08/85 14 12 29 SCALE 20000)

This computer-assisted reference map at scale 1:20 000 is among the first of a set of 108 being produced by Statistics Canada as part of a project undertaken by the Trillium Data Group to provide the Ontario Ministry of Health with a mapping and street indexing system for the Halton-Peel Central Ambulance Dispatch Services. The maps, compatible with base maps produced by the Ontario Ministry of Natural Resources, are to be included in a series of map books to be used by ambulance dispatchers and attendants.

segments long enough to accommodate the name. Possible overlap with other text strings was not considered; each name was placed with respect to the line feature to which it belongs. Interactive text repositioning was required, an average of $\frac{1}{4}$ hour per Census Tract map (Figure 3), and 6 hours per 5 km by 10 km map sheet (Figure 4).

The polygon text placement for municipality

reference maps (see capitalized names in Figure 5) was based on the minimum bounding rectangle of the polygon. The text string was broken at the appropriate blank(s) or hyphen(s) to give an overall shape approximating the shape of the rectangle, and placed by default at the polygon centroid. Thus, there was no checking for possible overlap with other text strings. This method worked well in sparse areas (i.e. large polygon size) and less well in denser areas.

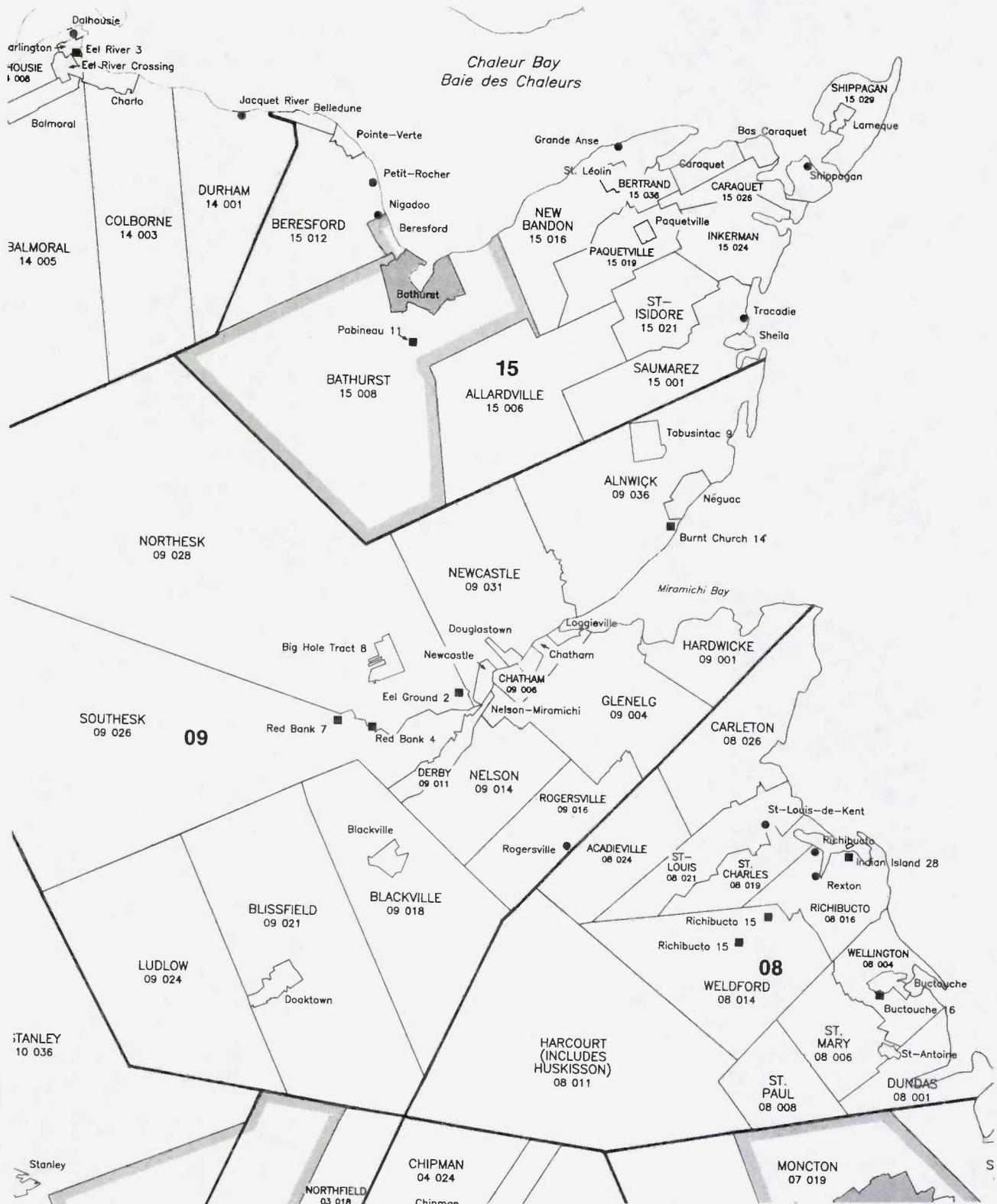


Figure 5 Part of the New Brunswick Reference Map of 1981 Census Divisions, Consolidated Subdivisions and Census Subdivisions, using the seven-digit Standard Geographical Classification (SGC) identification codes.

PANEL DISCUSSION AND COMMENTARY



Moderator:

Jean-Paul Drolet Chairman, Canadian Permanent Committee on Geographical Names, EMR

Roger L. Payne

Chief, Geographic Names Information Section, National Mapping Division, U.S. Geological Survey, Washington

Richard J. Darley

Senior Associate Chief, Cartography, National Geographic Society, Washington

Panelists:

John J.S. Thompson Assistant Director, Geographic Mapping, Geographical Services Division, EMR

Audience participants:

Michael R. Munro Head, Toponymy Section, Geographic Mapping, Geographical Services Division, EMR

Joel Yan

Chief, Methodology, Geocartographics Subdivision, Statistics Canada, Ottawa

Jeffrey R. Ball Cartographic Toponymist, Ontario Ministry of Natural Resources, Toronto

Norman Drummond

Associate Professor, Department of Geography, McGill University, Montréal

Gerald F. Holm Toponymist, Surveys and Mapping Branch, Department of Natural Resources, Winnipeg

Michael Smart

Executive Secretary, Ontario Geographic Names Board, Ministry of Natural Resources, Toronto

Jean Poirier Assistant to the President, Commission de toponymie du Québec, Québec

David Arthur

Secretary, Saskatchewan Geographic Names Board, Regina

Clément Nolette Graduate student, Department of Geography, Université Laval, Québec

Douglas W. Kirk

Hydrometric Systems Engineer, Water Survey of Canada, Water Resources Branch, Environment Canada, Ottawa

Donald Orth Chief, Branch of Geographic Names, National Mapping Division, U.S. Geological Survey and Executive Secretary, Washington

William C. Wonders

Professor, Department of Geography, University of Alberta, Edmonton

Marion O. Harris

Consultant, Bell Communications Research, Madison, N.J.

Léo La Brie

Translator, Translation Bureau, Secretary of State, Ottawa



A variety of topics and questions were raised for discussion. As far as possible, items have been grouped to assist the reader, and are presented under the following headings:

- (1) The contents and content limits of toponymic data bases
- (2) Coordination of geographical names data bases with other data bases and with digital cartography
- (3) Users and their demands for geographic names information
- (4) Particular problems in the automation of geographical names:
 - (a) odonyms
 - (b) accented characters

- (5) Pronunciation of geographical names, and concise gazetteers
- (6) Recommendations

- (1) THE CONTENTS AND CONTENT LIMITS OF TOPONYMIC DATA BASES

Mr. Thompson explained that the Surveys and Mapping Branch had made a conscious decision at the outset of the automation of the National Toponymic Data Base (NTDB) to create a toponymic system, rather than a comprehensive geographic information system. Attributes, such as popula-

tion, size, climatic conditions, average annual income, elevations, etc., that are not strictly toponymic in nature were not considered for inclusion. However, Mr. Thompson indicated that eventually toponymic material might be merged with other ongoing work into a broader geographic information system.

Mr. Orth reported that in the United States each geographic name is intricately tied to other data that can be identified with a particular landscape feature. As a result, the US approach has not been restrictive, but has been one of expansion to include information not already covered by other fields, such as automated cartography. Mr. Orth pointed out that there is, of course, a limit to the type of data they would include. For stream flow, for example, the approach would be reversed; the Geographic Names Information System (GNIS) would provide the basic information (or a subset), which could be used by the Water Resources Division of the USGS and merged with their files. Mr. Payne emphasized that GNIS is still an automated names system, rather than a geographic information system. Some elements or data categories (e.g. population) are now being added, as a result of user requests. In the case of population figures, the US Bureau of Census (which has membership on the USBGN) accepts the responsibility for the integrity of the data added to the system.

Mr. Darley explained that the National Geographic Society had developed an internal spelling checker, but found that it could not be cost effective unless other functions could be performed with the names data file. With population and other attributes added, it will be possible to automatically retrieve the bulk of the map type from the data file. This saving in person-hours will help offset the cost of building the system.

Mr. Holm reported on the progress with computerization of names records in Manitoba during the past 10 years. The VAX 11/750, purchased in 1983 is used with a DEC TE 16 tape drive and DECUS (DEC User-Society) software, which provides an editor, a spelling checker and a tape manipulation program. Nineteen terminals and two printers are now in use within the Branch; two terminals and one printer have international character capabilities. A tape of Manitoba records in the NTDB in Ottawa has been compared with the set of data used to produce the Manitoba Annual Directory. An error edit listing has thus been produced and will be available to correct NTDB records. Mr. Holm reported that working with Surveys and Mapping personnel at 615 Booth Street he had just made the first dial-up and outside access to Manitoba's computerized names system. Manitoba's system is now compatible with those of most provincial Surveys and Mapping Branches in Canada who have VAX or DEC PDP 11s. At present, Manitoba is using the system to prepare a cumulative supplement for the Manitoba volume of the "Gazetteer of Canada"; this is a quick process with a printed version available within a couple of hours.

Mr. Poirier reported that due to lack of funds the Commission de toponymie did not use the ONOMA system. It is now expected to have micro-computers and graphics in use for toponymic information in two or three service areas of the Commission.

Mr. Arthur expressed the interest of the Saskatchewan government in an automated toponymic data base. He hoped that within a few years funds might permit the acquisition of equipment and a hook up to the NTDB. In the meantime the old card record system would have to suffice.

Mr. Orth and Mr. Payne pointed out that in the US no states had automated data bases of geographic names



SYMPOSIUM PANELISTS:

(Left to right): Roger L. Payne, Donald Orth, Richard J. Darley, Gerald F. Holm

prior to the computerized GNIS. Subsequently, however, various states have established automated names data bases by incorporation of GNIS data in their own state systems. As far as geographic names and their applications are concerned, use of the USBCN decisions are incumbent upon all federal agencies.

(2) COORDINATION OF GEOGRAPHIC NAMES DATA BASES WITH OTHER DATA BASES AND WITH DIGITAL CARTOGRAPHY

Mr. Orth elaborated on some details of GNIS that had been developed to integrate it with the digital mapping program. For example, geographic coordinates have been recorded to the nearest second (with a tolerance of ± 5 seconds) and for streams the coordinates of both the source and the mouth have been recorded. Mr. Payne added that scale does not pose a problem. The base map series used for the digital mapping program is 1:24,000 or 1:25,000. The digitizing of names information was done from paper copies at the largest scale available at the time, i.e. 75% of Phase I information was digitized from 1:24,000 scale topographic maps; the remaining 25% was from 1:62,500 scale sheets. Mr. Payne explained that a research program had just been initiated to establish a link between GNIS and the digital cartographic data base; problems (including the question of digitizing from paper copy rather than original art work) will be identified during the initial phase of research.

Mr. Ball indicated that in Ontario digitizing for the provincial geographical referencing system was being started on a prototype basis. Position information that is put into the geographic names files and hence into the future automated data base must be compatible; some items, such as the differences between using UTM referencing and geographical coordinates, are posing questions.

Mr. Yan commented on the important effort being made in the US by the establishment of a federal digital cartography committee to coordinate all federal cartographic and geographic information systems activities. Mr. Thompson explained that representatives from various federal departments are starting to address the problems of incompatibility in the exchange of digital information among Canadian federal agencies. The committee, which had held its first meeting in Ottawa, plans to address the problems of the plethora of digital information and to take such steps as exchanging staff, and exchanging proposed standards pertaining specifically to geographic information and cartographic output. Hopefully the committee will also look at the treatment of toponyms, accents and standard definitions of generics.

With regard to links with digital cartography, Mr. Orth envisaged that in future individuals, rather than consulting paper maps, would be able to call up selected information on a CRT in their office, home or car. In order to meet demands of users selecting data fields for a specified area, the geographic names data base must be able to provide the reference information needed for this type of future mapping. Mr. Orth noted that, in fact, at the USGS the geographic names files are farther advanced than many other aspects of digital cartography.

Dr. Drummond asked about automated names placement and how a data bank of geographical names could be coordinated with a data bank of resources to get an accurate integration, for example, of names and geological units and their codes.

Mr. Thompson said that other than for examples such as type used on census mapping material (Statistics Canada) there is definite difficulty in sizing and placing names with proper orientation even for a single purpose, yet alone integrating with a second group of names information. Current cartographic standards may have to be changed; however, this process, is at present a low priority.

Mr. Orth felt this to be getting into cartographic problems and not necessarily tied in with names standardization or automation. Automatic names placement is in its infancy, but no doubt before too long there will be automated identification and placement of all elements on a map. He added that technically US geologists may not use unofficial geographic names in reports; geological names, although based on geographic ones, do not fall within GNIS data.

(3) USERS AND THEIR DEMANDS FOR GEOGRAPHIC NAMES INFORMATION

Mr. Payne commented on the users of GNIS data. He said that relatively complete records are kept of users of the system; in general terms the breakdown of users is in the order of 60% private industry, 35% government and 5% other. A complete cross-section of society is included, from oil companies and trucking firms to individuals pursuing genealogical research. Mr. Payne mentioned that 10-20 orders/week is common. The type of requested information varies; however, one of the most common is for all mountains within a certain state and over a particular elevation. He reported other types of data enquiries related to populated places over 6000 feet for a company distributing a particular product; information to plan delivery routes within a 100-mile radius of Houston; cemeteries within a particular radius; all lakes and reservoirs in a particular region. All these requests can be filled with a fraction of the time and cost involved in map searches to obtain the same data. Mr. Payne noted that the largest use made of GNIS data is probably unknown; orders for the purchase of digital tapes are constantly being filled through a clearing house. The information is dumped into the data base of the purchasers and subsequently applied to their own needs.

Dr. Wonders questioned Mr. Ball on the information base on which he had drawn the conclusions in his paper. The 17 responses appeared to assess government-user reactions to the survey. He wished to know whether there was any statistical basis for assessing the use of Ontario's geographical names data by non-government agencies, including companies, academic institutions, writers, private citizens, etc. Secondly, he felt that the survey results were certainly skewed when no weighting factor had been included for the actual number of enquiries from one source.

Mr. Ball replied that the 21 people contacted had been identified from other statistics as the primary and secondary data users; they formed the main target group for this particular study on use of various aspects of the written and graphic records.

Mr. Darley asked whether Ontario would be defining areal features as polygons, and rivers as linear features with many coordinates along their courses. Mr. Ball said that only a minimum number of coordinates would be recorded: traditionally the centroid for a polygon, and the mouth and source for rivers, plus the coordinates of the beginning and end on any particular map sheet.

He thought that the idea of putting in more coordinates so that the area could be defined and automatic type placement could be developed, was well beyond the capability of a toponymic system.

Dr. Drummond was interested in what feedback is obtained from resource departments, how these users select or eliminate names for their thematic maps and whether they have their own place name files, which might, for example, associate occurrences of rock types with particular geographical names.

Mr. Ball said that in Ontario, the Ontario Basic Mapping (OBM) Program establishes its own criteria for the use of names, and that other users in the province could do likewise. He indicated that usually geologists request that names be given to unnamed features, to clarify their reports. Mr. Smart added that OGNE records would supply information on official names, as opposed to unofficial names, perhaps used internally by survey parties. Mr. Ball explained that geographic nomenclature maps were being produced as an OBM project. Each map shows approved nomenclature to a certain date. New decisions in future will be incorporated manually onto the cronaflex, both on the map face and the names listing on the margin of the map. A copy of the updated base map could be sent to those on a select mailing list.

Mr. Holm spoke of supplementary names maps made available in November 1984, on a cost-recovery basis, to users in Manitoba provincial government departments. These 1:250 000 maps overprinted with up-to-date names data were aimed particularly at resource personnel in the northern part of the province. A copy was also made available for general use in the provincial map office in Winnipeg. Mr. Holm added that most enquiries did not deal with clutter of names information on maps, but rather with the many unnamed features. Data requests from resource companies can be met by providing a dump of the province's complete Annual Directory file, which can be done cheaply and at short notice. Unofficial names that have been used in the past, by explorers or geologists for example, may be in the records as variant names; no complete list of such names is available at present.

Mr. Thompson explained that Surveys and Mapping Branch has responsibility for the National Atlas of Canada as well as operation of the NTDB. One way of responding to problems of cartographic and toponymic clutter is to select a number of base maps upon which geographers can draw. Pre-formatted base maps could include light, medium or heavy drainage, town stamps, etc.; various menus could then be chosen to give appropriate combinations of information. This approach prevents the user having complete carte blanche, but instead introduces some rigour by providing consistency of product, while at the same time decreasing errors in compilation of items such as place names. Pressures to include more names to complement particular thematic requests are resisted, but the selection of names for base maps would aim to include principal populated places, climatic stations, etc.

Mr. Kirk asked several questions concerning the interactive supply of data, and the availability of user-friendly systems in Canada. He felt that data bases often ended up being modified clones of each other; he wondered if consideration had been given to using the same program for retrieving data from various provinces. In the US, with the user-friendly GNIS, Mr. Kirk wondered if the number of requests for information had diminished and whether particular users have access to different files.

Mr. Thompson replied that the NTDB is user-friendly to someone who is familiar and comfortable with the system; documentation is fairly lengthy and unfortunately with the current hardware and software features it is not uncommon to lose a whole file or an afternoon's work at an interactive terminal. Although it is nowhere near as large as the US system, it is perhaps more complex and not easily accessed from a home computer. As the provinces develop their own systems, each has different priorities with regard to financial and human resources. In Canada, the federal government could not wait for each province to have a geographic names data base in place and then develop and coordinate a central system. Instead, as mapmakers, Surveys and Mapping Branch needed to use the names made available from the CPCGN Secretariat, and had to move ahead.

Mr. Payne emphasized that "user-friendly" means different things to different people. Basically, at the USGS, the names office is research-oriented, rather than user-service oriented. From the initial lack of user-friendliness, the system since 1980 has become more and more friendly (with a menu-type system), with the user services being handled by about 20 information outlets, where the enquiry staff have had but a short training to learn the commands of the system. The difficult part is to understand the nature of the data in the system, in order to properly answer customers' questions; this is the area of training that is now being given attention. Mr. Orth added that in striving to make the system more user-friendly, codes have been eliminated as much as possible; looking up codes introduces a time factor and possible error factor. Mr. Payne commented that codes are retained for digital processing and are, therefore, used on magnetic tapes; but in response to user requests, clear type without codes is used for enquiries.

Mr. La Brie asked what steps would be necessary if a translator wished to know, for example, all names with a certain generic; how could this be done and at what cost?

Mr. Thompson requested that particular enquiries be addressed to the CPCGN Secretariat. System capability allows for call up of all generic codes by province and more specifically by parameters, such as MTS map sheet or geographic windows of latitude and longitude. Efforts are made to work on a cost recovery basis. For other government departments and agencies, if staff time is minimal, straight recovery costs are charged; only direct labour and systems costs are recovered, not supervisory or overhead costs. For example, the cost of a tape of all Manitoba names sent to the provincial government was several hundred dollars. Basically, the first request for a count of lakes, for instance, might cost a few hundred dollars; the second request for the same information would cost much less because the file would already exist.

Mr. Payne explained that costs for data from GNIS depend on the extent of the area of search. A search for "ponds" in one state would cost between \$1.50 and \$10, whereas a search of the whole US set of files (more than 2 million entries) would cost \$50-60 US. Gradually over the last five years customer enquiries have increasingly been referred to the National Cartographic Information Center, which has interactive access to GNIS. The system was originally designed for this type of enquiry and does it very well. Records are kept of the requests received. Mr. Orth added that a list of every kind of generic for any definable polygon could be provided. This information can also be plotted on maps,

as has been done with the distribution of various generics in the US.

Mr. Poirier explained that as recommended by the United Nations, the Commission de toponymie makes a distinction between the generic, as "crique" in Crique Honorius, and the "entité", which is the broader classification of generic terminology, in this case "ruisseau".

(4) PARTICULAR PROBLEMS IN THE AUTOMATION OF GEOGRAPHICAL NAMES

(a) Odonyms

Mr. Poirier reported on odonyms (street names) in Quebec. In 1979 the "Guide toponymique" was produced to provide rules for writing geographical names, including street names. Currently the Commission de toponymie shares with the municipalities the responsibility for officializing street names, for which the Commission estimates a total of some 100,000 to be formalized. Recently the Commission began entering street names into the data base and some of them have already been submitted to the Commission for approval. It is hoped that in a year's time it will be possible to retrieve on the computer odonyms by municipality, and in two years to be able to list all toponyms in a particular municipality. At present the Commission is in a transition period, experimenting with the use of micro-computers. Recently, for greater precision, coordinates have been recorded for both ends of a street. This procedure is not yet followed for watercourses as there are more problems in identifying exactly the beginning of a river in a swamp or basin. Although autoroutes may be identified by number, only their names are entered in the official records, and the Commission will try to indicate geographical coordinates within the various municipalities through which the autoroute passes.

Mr. Orth indicated that the entry of odonyms into GNIS would be a Phase III activity. For two reasons these data have not previously been entered. (a) Until now it has been very difficult to collect this information consistently across the country; the efforts of the US Bureau of Census and US Postal Service are now changing this situation. (b) It has not been resolved how to identify, within the system, the extent of the features. Now it is hoped to tie into the digital cartography program, once a start is made on digitizing streets and roads; perhaps then, particular points can be identified along the feature and a name applied to the string of data points.

(b) Accented characters

Mr. Darley reported that so far the National Geographic Society is looking at automatic type retrieval, not automatic type placement, from a system that will have world-wide accents, will display them on a screen and will be able to print them on a small printer. With correct spelling and accents, tapes generated for typesetting will produce more accurate results. Using their large IBM mainframe computer, and a graphics package purchased from IBM but slightly modified for an on-shelf terminal with standard keyboard, National Geographic has built alphabets to include all letters that need accents. The information can be displayed on

the screen and the type setting equipment instructed which characters require accents, in either upper or lower case. Mr. Darley indicated that this standard IBM graphics package has, in fact, been adapted to the National Geographic Society's use and is working beautifully.

The first names to be put into the National Geographic's system will be the names from their World Atlas; information will be updated as it is entered, using official spelling as it should appear on maps. One problem is how the magazine may want to use the data in publishing for an English-speaking audience. To cater for this, names cataloguing must take into account map, magazine or book preference. In the magazine there is no wish to have accents, but on maps accents would be used, depending on the transliteration system. Both sets of names must be entered and labelled so that the researcher can extract whichever version of the name is required. Mr. Darley stressed that the characters are created out of a matrix on the screen. When the graphics package was received, National Geographic had to design and enter every letter and every configuration of possible accents into the system to build up the file themselves.

Mr. Orth explained that the USGS will soon be facing this question of accents. At present they can handle diacritical marks found in French and Spanish, but are currently moving into recording Indian names, specifically Navajo and Apache with their own modified Roman alphabets, using slashes, letter combinations and apostrophes above and below letters. As this will be a very important area, Mr. Orth hoped that some commercial companies will make a break through in providing keyboards to handle a great variety of less familiar symbols.

Mr. Yan expressed the wish that the question of designing and using special characters be followed up.

(5) PRONUNCIATION OF GEOGRAPHICAL NAMES, AND CONCISE GAZETTEERS

Mr. La Brie was interested to know whether pronunciation would be provided for the 50,000 names to be included in the "Concise Gazetteer of the United States of America".

Mr. Orth responded that Canada is ahead of the US in pronunciation standards. In many ways a "standard" pronunciation is very difficult to ascertain; for example, Chicago has at least 15 local pronunciations. He reported that the concise gazetteer was developed in response to a recommendation of the UN Group of Experts on Geographic Names, when a national gazetteer program for the whole world was planned and countries assigned total numbers of entries on the basis of size and population. Libraries in the US also favoured one distilled volume of populated places and physical features as a quick reference, rather than referring to individual state gazetteers. One base for names was the "National Atlas of the United States of America", which contained 40-50,000 geographic names, predominantly populated places. To this were added names of major physical features to give a total of about 50,000. No pronunciation guide is being considered at this point.

Mr. La Brie commented that Austria is the only

EXTRAITS DU DISCOURS D'OUVERTURE

DE

JEAN-PAUL DROLET*

Il me fait extrêmement plaisir de vous souhaiter la bienvenue au deuxième Colloque sur l'automatisation des noms géographiques. La grande participation que l'on peut remarquer ici aujourd'hui reflète l'intérêt marqué que les comités de toponymie et bureaux de cartographie du Canada et des États-Unis portent à l'établissement d'un système de classement fiable et pratique des renseignements relatifs aux noms géographiques et à l'amélioration des échanges d'information par l'informatique.

En organisant ce colloque, le Comité permanent canadien des noms géographiques accompli, j'en suis sûr, l'une des fonctions importantes pour lesquelles il a été créé en 1897: uniformiser les noms géographiques utilisés au Canada et réunir les divers représentants d'organismes provinciaux et fédéraux de toponymie pour résoudre les problèmes communs.

D'abord, j'aimerais remercier ceux qui ont préparé des exposés pour ce colloque: John Thompson de la Direction des levés et de la cartographie du gouvernement fédéral; Jean Poirier de la Commission de toponymie du Québec; Jeffrey Ball du Ministry of Natural Resources de l'Ontario; Clément Nolette du Département de l'informatique de l'Université Laval et Roger Payne de la Domestic Names Section du United States Geographic Board. De plus, d'autres personnes ont accepté de participer avec les conférenciers à la discussion spéciale de cet après-midi: Richard Darley du National Geographic Magazine, Washington; Donald Orth, United States Geographic Board; et Gerald Holm du Manitoba Department of Natural Resources. Participent également à notre réunion, Randall Detoro de la Nicholls State University à Thibodaux, Louisiana. Il est le directeur de la United States Placename Survey.

Le programme comporte cinq exposés officiels, suivis d'une discussion en groupe; il pourra y avoir ensuite une période de questions. J'aimerais que le programme soit aussi souple que possible de façon à ce que d'autres personnes qui travaillent elles aussi à la mise au point de systèmes informatiques en matière toponymique puissent apporter leur contribution. Colin Bromfield du Service hydrographique canadien, John Carey du Bureau des traductions et Joel Yan de Statistique Canada ont préparé quelques documents portant sur les programmes informatisés utilisés dans leurs bureaux respectifs et sont prêts à faire de brefs exposés.

* Monsieur Jean-Paul Drolet, Président, Comité permanent canadien des noms géographiques.

Il y a cinq ans, presque jour pour jour se tenait à Ottawa le premier colloque sur l'automatisation des noms géographiques. A cette occasion, nous avons entendu d'excellents exposés sur la théorie et l'élaboration des systèmes informatiques, sur leur implantation et sur la nécessité pour les utilisateurs de disposer de renseignements fiables et sans ambiguïté sur les noms géographiques.

Au colloque de 1980, j'ai soulevé un certain nombre de questions fondamentales sur l'automatisation des renseignements relatifs aux noms géographiques; ces questions et plusieurs autres qui ont été abordées au cours de ce colloque méritent qu'on les rappelle ce matin.

1. De quelle façon peut-on utiliser la technologie de l'informatique moderne pour enregistrer des renseignements de nature toponymique?
2. De quelle façon la technologie de l'informatique peut-elle accélérer l'accès à l'information toponymique?
3. Comment peut-on s'assurer qu'il y aura compatibilité entre les systèmes pour permettre l'échange efficace des renseignements essentiels?
4. Quelles sont les limites, en matière de contenu, d'une base de données toponymiques?
5. Comment peut-on améliorer la qualité et la disponibilité des renseignements toponymiques compte tenu des restrictions budgétaires qui empêchent d'accroître le nombre de personnes affectées à l'enregistrement et à la manipulation des données toponymiques?
6. Par quel moyen peut-on protéger les renseignements toponymiques de toute modification, une fois qu'ils ont été autorisés par l'autorité compétente?
7. A quoi s'attendent les utilisateurs?
8. Dans quelle mesure peut-on s'attendre à ce que les utilisateurs remboursent les coûts d'élaboration et de gestion du système?
9. A quel moment conviendrait-il d'abandonner les fiches classiques sur papier et de se fier totalement aux systèmes informatiques?

Il ne s'agit là que de quelques-unes des questions auxquelles il faut répondre.

LA BASE NATIONALE DE DONNÉES TOPONYMIQUES - COMPTE RENDU DES TRAVAUX

J.J.S. Thompson* et M.R. Munro**

[Translation]

RÉSUMÉ. La base nationale de données toponymiques (BNDT) est une base de données numériques et analogiques intégrée, spécialisée dans les noms géographiques, et gérée par Énergie, Mines et Ressources Canada. Depuis son automatisation en 1978, cette base de données a subi des modifications structurales importantes, notamment la réduction du nombre de fichiers, qui a facilité les transactions. Une étude est actuellement en cours pour évaluer les exigences futures en matériel et en logiciel de la BNDT qui devra faire face à une demande grandissante et à une expansion de son contenu et s'engager dans une répartition accrue de ses services.

INTRODUCTION

La base nationale de données toponymiques (BNDT) est une base de données numériques et analogiques intégrée, spécialisée dans les noms géographiques, contenant les toponymes approuvés par le Comité permanent canadien des noms géographiques (CPCNG). La Section de toponymie de la Direction des levés et de la cartographie est chargée du fonctionnement et de la mise à jour de cette base de données et, par l'intermédiaire des bureaux du Secrétariat du CPCNG, de l'accès des données toponymiques au gouvernement fédéral et ses organismes ainsi qu'au grand public. Il y a un certain temps, par suite de l'augmentation continue du nombre de fiches toponymiques de type analogique et du besoin constant d'accroître l'efficacité des méthodes d'acquisition, de stockage et d'extraction des données toponymiques, il est devenu évident qu'il faudrait réviser les méthodes manuelles classiques utilisées par la Section de toponymie pour mettre à jour et faire fonctionner le fichier des noms géographiques.

AUTOMATISATION INITIALE

En 1977, la Division des services de géographie a analysé les services fournis par la Section de toponymie et la qualité des fiches toponymiques utilisées à l'époque, en particulier, les fiches-index et les documents classés dans des chemises. L'analyse a révélé que compte tenu de la grande dimension de la base de données (évaluée en 1977 à 350 000 noms enregistrés sur fiches-index et à 650 000 dossiers contenus dans des chemises de classement) et d'un taux annuel de 7 % d'ajouts et de modifications, le système d'information toponymique nécessitait la mise en place d'une technologie plus moderne de manière à satisfaire aux exigences des clients. Pour améliorer le stockage et l'extraction des données et pour rendre la base de données toponymiques plus compatible avec les éléments automatisés de la cartographie, il a été décidé d'informatiser les fichiers toponymiques.

Suite à cette évaluation, un bureau d'experts-conseils en logiciel a été engagé à contrat pour faire

des recommandations sur l'automatisation de la BNDT. Le rapport présenté par ce bureau recommandait l'achat d'un progiciel de gestion de base de données fabriqué sur mesure et fonctionnant à partir d'un mini-ordinateur. Cependant, les cadres de la Division, après s'être minutieusement penchés sur la question, ont finalement conclu qu'il fallait d'abord évaluer plusieurs des nombreux progiciels qui étaient commercialement disponibles et qui pouvaient offrir des possibilités intéressantes en ce qui a trait à l'exploitation de la base de données toponymiques. Le choix a finalement porté sur le progiciel DATABOSS/2. Ce système, qui utilise le langage de programmation Basic-Plus permet le traitement en direct ou par lots et des entrées de données provenant de tout fichier de l'ASCII. De plus, il fonctionne sur des mini-ordinateurs DEC PDP et un système d'exploitation RSTS/E. La compatibilité entre le DEC PDP et le RSTS/E a été un critère de sélection particulièrement important, étant donné que la Division disposait à l'époque d'un ordinateur DEC PDP 11/45.

Comme projet pilote, il a été décidé de produire une nouvelle édition du Répertoire géographique du Canada: Territoires du Nord-Ouest en utilisant le nouveau système informatique. Les renseignements toponymiques contenus sur fiches-index ont été transcrits sur des formules de codage informatique avant d'être enregistrés sur bande magnétique par une société de traitement de textes. Pendant ce temps, IAS Computers Ltd. était engagé à contrat pour fournir et installer le progiciel DATABOSS/2 et le système d'exploitation RSTS/E sur un ordinateur PDP 11/45 de la Direction. Le système a été mis en communication directe conformément aux directives détaillées de stockage et d'extraction, après quoi les données toponymiques des Territoires du Nord-Ouest contenues sur bande magnétique ont été chargées dans le système. Les renseignements ont été par la suite corrigés en direct et c'est le nouveau système qui a produit la version finale sur bande magnétique. Le Répertoire géographique du Canada: Territoires du Nord-Ouest, 1980 fut donc le premier répertoire réalisé par la base nationale de données toponymiques informatisée. Depuis, cinq autres répertoires ont été produits dans cette série, notamment ceux du Manitoba, du Yukon, de Terre-Neuve et de la

* John J.S. Thompson, Directeur adjoint, Cartographie géographique, Division des services de géographie, Direction des levés et de la cartographie, EMR.

* M.R. Munro, Chef, Section de toponymie, Cartographie géographique, Division des services de géographie, Direction des levés et de la cartographie, EMR.

Colombie-Britannique. Celui de la Saskatchewan est actuellement en cours; il devrait être publié pour l'été 85. La BNDT fait l'objet de mises à jour et de modifications constantes de manière à ce qu'elle reflète bien les dernières décisions provinciales en matière de toponymie.

Le progiciel DATABOSS/2 est composé d'un seul fichier qui, bien que pouvant contenir un très grand nombre de fiches, impose un certain nombre de contraintes au système toponymique, entre autres la longueur des fiches. Pour optimiser le rendement de ce système commercial, on a procédé à des travaux de programmation sur place visant à transformer ce système à fichier unique en un système à sept fichiers. Ces fichiers sont les suivants: le fichier des données toponymiques (incluant un identificateur unique, des champs pour le nom de l'élément, le générique, la province, la date de la décision, le statut, la latitude, la longitude, les champs relatifs à la production du répertoire et la grosseur des caractères); le fichier des données relatives aux cartes; le fichier des données relatives aux renvois; trois fichiers contenant des données sur l'emplacement du nom; et le fichier narratif sur l'origine des noms. (Figure 1)

Les fonctions dont disposent les utilisateurs du système sont la création et l'affichage de fichiers ainsi que l'insertion, l'annulation, l'affichage et la correction de fiches. En outre, trois autres programmes ont été mis au point, dont deux ont été spécialement conçus pour l'entrée de données (le programme d'entrée

de données en mode interactif et le programme d'entrée de masse); le troisième fichier est appelé le producteur de rapports.

En général, on peut affirmer que les essais initiaux d'automatisation ont permis de créer un système plus fonctionnel que le système original en offrant plus de possibilités que jamais en ce qui a trait au traitement des données relatives aux noms géographiques. Il faut néanmoins préciser que nombre des fonctions énumérées se sont accomplies avec plus de lenteur et de lourdeur que l'on ne s'y attendait de la part d'un système informatisé. Il était également clair que compte tenu des limites du matériel et du logiciel, le rendement continuerait à se détériorer à mesure que le contenu de la base de données augmenterait. C'est pourquoi, il a fallu effectuer une certaine reprogrammation du système de façon à tenter de minimiser les effets de ces contraintes.

Cette reprogrammation a entraîné la conservation de toutes les fonctions utiles qui existaient dans le système original ainsi qu'une nette amélioration de la manipulation des fiches du point de vue de l'utilisateur. De plus, elle a entraîné la création de plusieurs nouvelles fonctions qui ont facilité l'utilisation du système. Les nouveaux programmes ont permis d'optimiser la capacité résiduelle du système (fondé sur la structure en sept fichiers). Malheureusement, l'implantation de ces nouveaux programmes limitait l'expansion éventuelle du système, de sorte que, malgré ces diverses améliora-

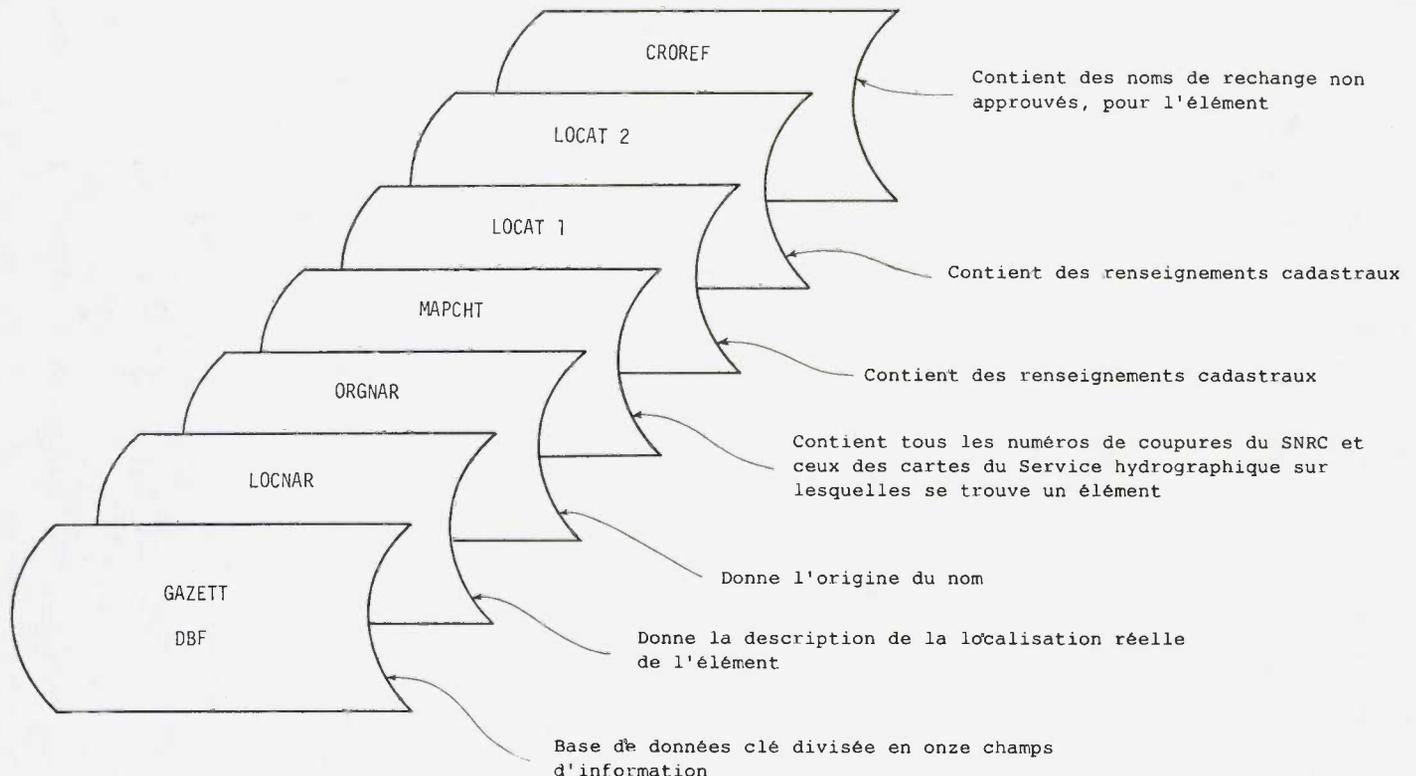


Figure 1 Structure originale de la base nationale de données toponymiques composée de sept bases de données

tions, les utilisateurs du système ont continué à se sentir lésés par la structure en sept fichiers, surtout lorsqu'ils devaient consulter chacun des sept fichiers individuellement pour accomplir certaines opérations.

CONFIGURATION ACTUELLE

Pour faire face à l'expansion croissante de la base de données et à la demande accrue pour obtenir un temps de réponse plus court de la part du système, une nouvelle configuration impliquant deux modifications importantes de l'architecture du système a été mise en place:

- a) La transformation des sept fichiers originaux (figure 2) en quatre fichiers plus grands: les fichiers Texte, Carte, Narration et Répertoire (figure 3);
- b) l'établissement de 14 sous-fichiers (dix pour les provinces, deux pour les territoires et un pour les entités sous-marines ainsi qu'un autre pour les eaux internationales).

Ce rajustement de la structure des fichiers de la base des données permet d'avoir accès plus rapidement aux données et de s'assurer que la base pourra contenir le nombre projeté de noms géographiques. La gestion du système a également subi des améliorations visant à faciliter la mise à jour et la vérification des fiches. Le logiciel, créé sur commande, prévoit la transparence de la séparation entre les quatre fichiers de la base des données ce qui permet à l'utilisateur d'obtenir les données relatives à chaque fiche au cours d'une seule

consultation. De cette façon, il est possible d'effectuer beaucoup plus rapidement des extractions, des corrections et des ajouts ou annulations de fiches.

Outre ces modifications, un certain nombre de nouveaux programmes ont été rédigés. En plus de fournir les fonctions qui étaient disponibles dans le système original et celles qui ont été ajoutées par suite de la reprogrammation subséquente, ces programmes permettent de rationaliser les mécanismes impliqués et d'offrir quelques services entièrement nouveaux et très utiles.

LES PROGRAMMES FABRIQUÉS SUR COMMANDE SONT, A CE JOUR PAR TITRE D'ACTIVITÉ

- a) L'entrée de données sur les sources

Ce programme, tout comme le Programme original d'entrée de données en mode interactif, est spécialement conçu pour entrer de grandes quantités de nouvelles fiches dans la base de données de la région appropriée. L'utilisateur peut utiliser ce programme pour entrer des fiches dans le fichier source et effectuer des annulations, des corrections et des impressions de données qui sont en cours d'introduction. Le gestionnaire de la base de données peut ensuite autoriser de mettre fin à l'opération et de procéder à la fusion des renseignements de façon à les entrer dans la base de données principale au moyen d'une insertion de masse.

- b) Les transactions en direct

Grâce à ce programme, il est possible d'effectuer des transactions en direct avec les données contenues

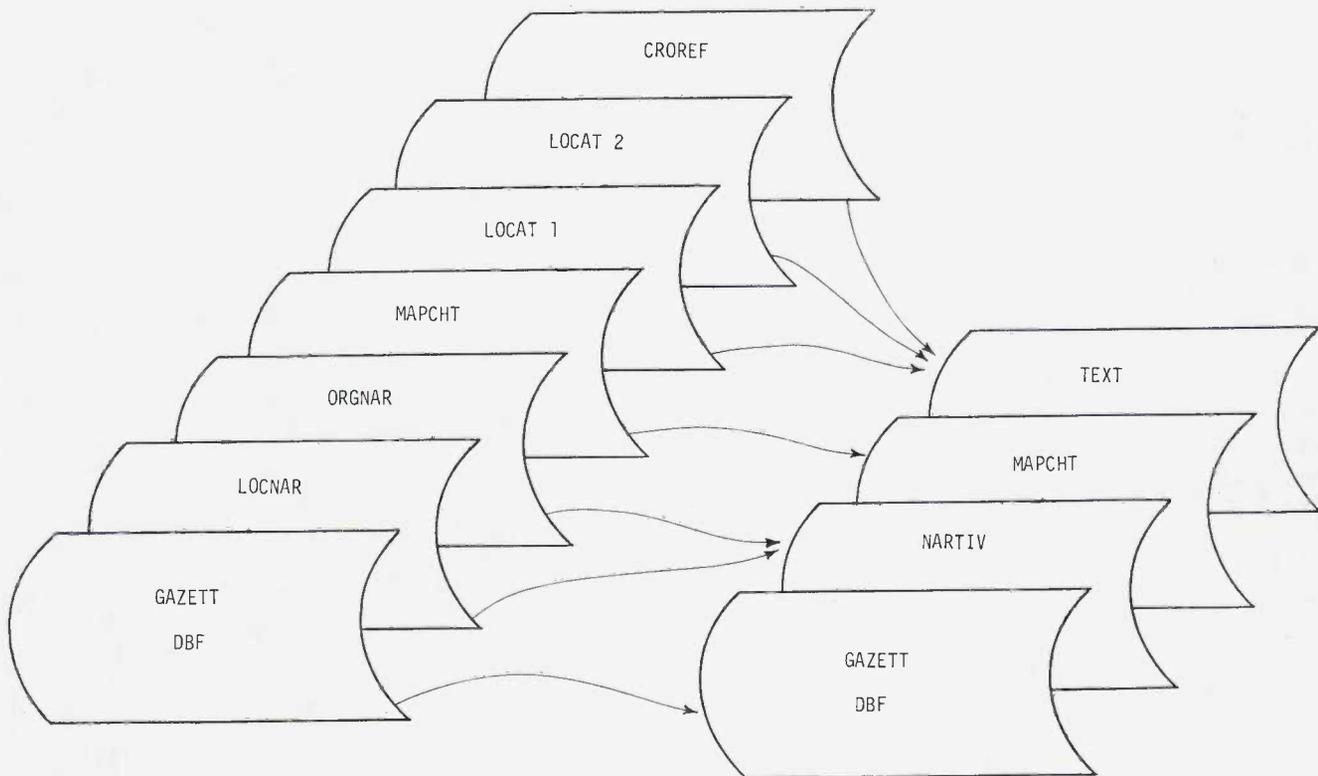
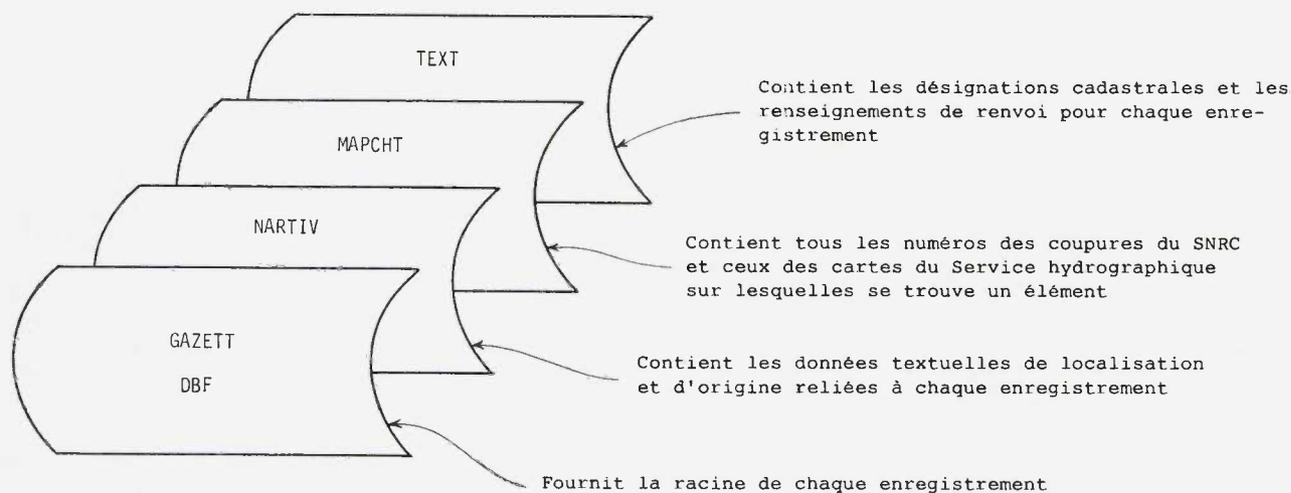


Figure 2 Regroupement des sept fichiers originaux de la BNDT en quatre fichiers révisés



(Chaque fichier est à son tour subdivisé en 14 bases de données régionales)

Figure 3 Structure actuelle de la base nationale de données toponymiques

dans la base de données principale. Ces transactions incluent les opérations d'annulation et de mise à jour ainsi que des mécanismes permettant de récupérer les transactions advenant la perte de données. Comme dans le cas du Programme d'entrée de données sur les sources, le gestionnaire de la base de données doit autoriser la fusion de ces transactions.

c) Les rapports de listes de noms

Ce programme a été conçu spécialement pour le personnel de la Section de toponymie qui est chargé de préparer les listes de noms qui doivent être imprimées conformément aux besoins des cartographes de la Direction. Un certain nombre d'opérations devaient être incluses dans ce programme. Parmi les tâches accomplies, mentionnons: la création d'un fichier de génériques; la mise sur pied d'un rapport de listes de noms; l'impression du rapport de la liste de noms; la création d'un fichier de contrôle des formats de la liste de noms qui fournit au personnel la possibilité de définir et de corriger en mode interactif les données toponymiques qui doivent être incluses dans le rapport; et la production du rapport révisé et la conservation du rapport final sur bande magnétique en vue d'une régénération dans le cadre d'une révision future de la feuille de carte concernée. Le fait que ces listes de noms soient emmagasinées sous forme numérique permettra une économie de temps énorme au moment de réviser les feuilles de carte dans l'avenir. (Figure 4)

d) La production des répertoires géographiques

La Section de toponymie étant chargée de produire la série des Répertoires géographiques du Canada, deux programmes ont été établis à cet effet. Le premier qui s'appelle le programme de rapport des répertoires géographiques permet au gestionnaire de la base de données de produire un imprimé des renseignements dont le format est conforme à celui qui doit être utilisé dans le ré-

pertoire donné. Cet imprimé est vérifié en fonction de la documentation originale et les corrections sont apportées dans la base de données principale. Pour faciliter ce travail, peu de modifications de programmation ont été apportées au système de sorte que le nombre d'occurrences de certains éléments (par ex. les génériques) peut être comté ou les noms des entités d'un certain type peuvent être imprimés. (Les génériques dont il est question sont les éléments génériques des toponymes). Le second programme de cette série est celui qui est utilisé pour créer une bande magnétique qui servira à la production de répertoires. Ce programme permet de produire une bande magnétique encodée qui peut être envoyée à une maison de composition qui produira l'original du répertoire requis par les imprimeurs. Cette bande contient tous les noms géographiques actuels approuvés ainsi que tous les noms approuvés anciennement mais qui ont subi des modifications ou des annulations.

e) Les travaux sur place et la liste de décision

Ce programme a été conçu pour permettre la création d'un fichier temporaire dans lequel les toponymes pourront être incorporés, corrigés, annulés, imprimés et, avec l'autorisation du gestionnaire de la base de données, transférés au fichier de la liste de décision. Le gestionnaire de la base de données peut autoriser la fusion de ce fichier avec la base de données principale.

f) L'extraction de renseignements généraux

Comme son nom l'indique, ce programme permet à l'utilisateur de demander l'impression ou l'affichage de tout toponyme ou groupe de toponymes contenus dans la base de données principale. En outre, il est possible de faire des extractions en utilisant les logiques "ET" ou "OU" par lesquelles l'utilisateur peut limiter sa demande à certains champs. Par exemple, il est possible de demander au système d'afficher toutes les entités du Nouveau-Brunswick qui contiennent le générique "rivière" ET qui contiennent le code d'état "A1" (toponyme approuvé).

```
Map Sheet / Carte: 62 K/12 SPY HILL SASKATCHEWAN MANITOBA
Project # / Projet #: 02-1916
Sheet Title Block / Titre de la Feuille:

SPY HILL
SASKATCHEWAN MANITOBA

----- PROVINCIAL BOUNDARIES / FRONTIERES PROVINCIALES

Saskatchewan F-3
Manitoba F-3

No. Name / Nom Location
-----
----- PLACE NAMES / NOMS DE LIEU

1 Yarbo A-3
2 Cutarm B-3
3 Gerald C-3
4 Spy Hill D-2
5 Hazel Cliffe A-1
6 Tantalion B-C-1
7 Welby F-1

----- LAKES, BAYS ETC. / LACS, BAIES ETC.

8 Miller Lake D-2
9 Perrin Lakes E-2
```

Figure 4 Exemple d'une liste de noms automatisés produite par ordinateur - BNDT

Il serait par la suite possible de demander au système d'afficher tous les noms comportant le générique "rivière" OU toutes les entités d'une région particulière comportant le générique "lac". La zone de recherche peut par la suite être limitée selon divers paramètres, notamment les données de latitude et de longitude pour délimiter une région géographique.

g) La programmation de la gestion de la base de données
Plusieurs programmes contribuent à rendre la gestion

du système plus efficace.

Mentionnons, par exemple:

- 1) le programme attribuant le numéro de carte du SNRC à toutes les fiches contenant des coordonnées de latitude et de longitude comprises à l'intérieur de la zone couverte par la carte;
- 2) le petit programme pour faciliter l'adoption des programmes existants;

- 3) le programme permettant au gestionnaire de la base de données de faire exécuter à des heures déterminées des travaux en traitement par lot;
 - 4) le programme permettant de modifier les priorités entre différents travaux qui sont exécutés en même temps;
 - 5) le programme permettant l'interruption de programmes en cours d'impression sans avoir à refaire les commandes déjà données;
 - 6) le programme permettant de faire des recherches en mode séparé;
 - 7) le programme permettant d'imprimer des travaux dans un ordre prescrit sans que l'opérateur n'ait à intervenir;
 - 8) le programme permettant de mettre fin aux transactions et de leur donner suite à distance.
- h) Fichier de description des génériques

Ce programme fait partie intégrante de nombreux autres programmes utilisés dans le système du fait qu'il peut être relié à un certain nombre d'autres programmes et en particulier au Programme de rapport des répertoires géographiques.

SÉCURITÉ

La protection de la base de données continue d'être assurée grâce à l'utilisation du numéro de compte de l'utilisateur, du code de protection et du mot de passe. De cette façon, le gestionnaire de la base des données peut contrôler l'accès à la base de données; de plus, la reprogrammation du système a contribué au resserrement du contrôle. L'obligation d'obtenir l'autorisation du gestionnaire de la base de données diminue sensiblement les chances que des modifications ou des annulations non approuvées soient apportées à la BNDT.

Des copies de secours sur bande magnétique sont effectuées sur une base quotidienne et hebdomadaire et une copie de secours des données traitées par le système est réalisée toutes les deux semaines. Ces copies de secours sont conservées en lieu sûr de façon à prévenir tout événement catastrophique ou tout acte de vandalisme.

MATÉRIEL INFORMATIQUE DE LA BNDT

Le matériel utilisé par la BNDT s'est quelque peu modifié depuis la mise en place originale du système. Les différences principales se résument à un accroissement de la capacité de stockage des données et à l'acquisition d'une imprimante rapide. De nouveaux terminaux ont remplacé la plupart des anciens tout en gardant les mêmes caractéristiques. Les composantes actuelles de ce système sont: (Figure 5)

- 1) un ordinateur DEC PDP 11/45 doté d'une mémoire principale de 256 Ko;
- 2) quinze chargeurs de disques;
- 3) un dérouleur de bande à double densité TE16 (800/160 BPI, 45 IPS);
- 4) cinq entraînements de disques RP06, d'une capacité

individuelle de 176 méga-octets;

- 5) deux multiplexeurs DZ-11 (pouvant être reliés à plus de 16 terminaux, y compris le soutien par composition);
- 6) une imprimante rapide General Electric 3400;
- 7) trois terminaux Volker-Craig 44;
- 8) un terminal de traitement de textes Volker-Craig 414;
- 9) deux terminaux Cybernex XLA100F;
- 10) cinq terminaux Cybernex XLA84F;
- 11) un VT100 de Digital Equipment Corporation;
- 12) une imprimante haute qualité Multi-Writer;
- 13) un terminal (console) à copie-papier LA34.

L'AVENIR

Une analyse actuellement en cours vise à déterminer dans quelle direction devraient s'engager les perfectionnements informatiques de la Base nationale de données toponymiques. On prévoit que la Direction des levés et de la cartographie aura besoin, avant 1987, d'un ordinateur plus moderne pour sa BNDT. L'ordinateur actuel est jugé archaïque selon les normes de l'industrie. Son temps de réponse est extrêmement lent, les pannes sont nombreuses et son utilisation est limitée à un nombre d'utilisateurs relativement restreint. Si la BNDT doit jouer un rôle important en cartographie automatique, ce problème devra être résolu.

La nécessité d'acquérir un nouveau logiciel est également à l'étude. Le logiciel actuel a été adopté et reprogrammé de telle façon que toutes ses possibilités de programmation sont déjà utilisées. Bien que l'acquisition du progiciel DATABOSS/2 ait contribué, il y a huit ans, à informatiser la BNDT à un faible coût et d'une façon relativement simple, il est de plus en plus évident que la forte demande en données toponymiques numériques conjuguée à une diminution des ressources humaines affectées au traitement manuel des demandes pour les noms géographiques devra entraîner un développement important du système.

Tout perfectionnement de la BNDT devra viser à assurer une accessibilité maximale de la base des données au plus grand nombre d'utilisateurs possible. D'ici à plusieurs années, on espère que la BNDT deviendra graduellement une base de données réparties à laquelle pourront avoir accès les autorités provinciales en matière de toponymie même si ce n'était qu'au niveau de la consultation. La collaboration entre le personnel de la BNDT, le Secrétariat du CPCNG et les autorités provinciales concernées s'intensifie en matière de production automatisée des répertoires géographiques et d'échange réciproque de renseignements toponymiques numériques.

De plus en plus, la Section de toponymie reçoit des demandes d'organismes provinciaux et fédéraux et du secteur privé qui veulent obtenir des listes toponymiques informatisées et des données numériques provenant de certains fichiers. C'est pourquoi, pour mieux répondre aux demandes des utilisateurs, des bandes magnétiques à format normalisé sont en train d'être mises au point.

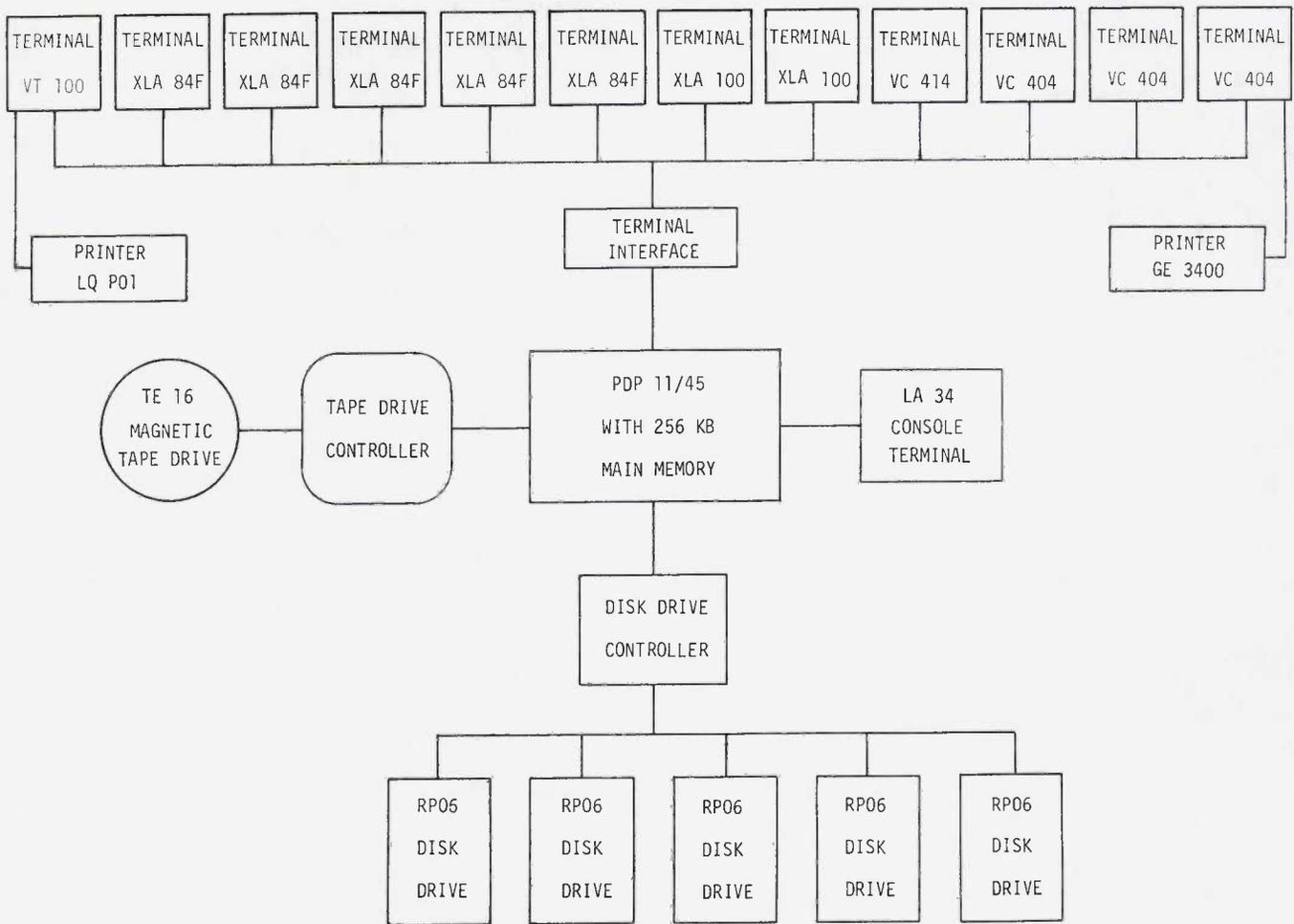


Figure 5 Matériel de la BNDT

Comme nous l'avons déjà mentionné, le fonctionnement de l'ordinateur actuel ne cessant de se détériorer, il a fallu faire une analyse des exigences auxquelles doivent répondre le logiciel et le matériel des systèmes actuel et futur de la BNDT. Parmi les exigences à satisfaire, mentionnons principalement l'extraction rapide et exacte des noms géographiques dans le cadre d'établissement de cartes à partir de données numériques et analogiques, en particulier les cartes de la série du système national de référence cartographique à diverses échelles, et la Carte internationale du monde, série de cartes à l'échelle de 1/1 000 000.

Notre principal objectif à cet égard est de fournir à tout cartographe qui en fait la demande quelle que soit la carte donnée, des renseignements toponymiques numériques à jour et exacts qui soient présentés selon une grosseur et un style de typographie appropriés.

CONCLUSION

Des progrès notables ont été accomplis pour automatiser la Base nationale de données toponymiques.

Les employés affectés à cette base de données, dont plusieurs n'avaient aucune expérience des ordinateurs et de l'automatisation, ont réagi positivement et de façon constructive à ce programme. Les répertoires toponymiques des provinces sont actuellement produits par un nombre beaucoup moins élevé de personnes et la réalisation de chaque répertoire permet de nettes améliorations dans la mise à jour des renseignements de la base de données.

Les perfectionnements apportés au système d'exploitation visant à rendre transparentes un certain nombre de transactions contribuent à rendre le système facilement utilisable. Une routine éprouvée de vérification et de contre vérification, de traitement de secours et d'analyse retrospective assure l'intégrité et la sécurité du système. Cependant, compte tenu de l'augmentation des demandes en matière de renseignements toponymiques numériques, de l'expansion continue de la base de données et du vieillissement rapide du matériel informatique, il faut dès maintenant que le matériel et le logiciel du système soient réévalués de façon à permettre à la BNDT de traverser la prochaine décennie dans problèmes.

Nous sommes conscients du potentiel énorme que

représente l'interface direct de la BNDT avec les extrants qui peuvent être facilement utilisés dans bien des produits cartographiques et géographiques et nous entrevoyons l'avenir avec enthousiasme en ce qui concerne le perfectionnement de la BNDT, avenir au cours

duquel on assistera à un échange accru de renseignements toponymiques numériques avec des utilisateurs intéressés et à l'établissement d'une base nationale de données toponymiques réparties, facilement accessible.

BASE DE DONNÉES TOPONYMIQUES DE L'ONTARIO: COMPTE RENDU D'UNE ÉTUDE DES BESOINS DES UTILISATEURS

Jeffrey R. Ball*

[Translation]

RÉSUMÉ. L'Ontario procède actuellement à une refonte de son système d'information toponymique pour l'adapter au système de références géographiques de la province, et aussi pour en faire une base de données utile sur les noms géographiques approuvés et enregistrés. La première phase de ce programme de refonte est déjà en cours; elle consiste à établir une série d'"index graphiques" dans lesquels sera répertoriée la nomenclature géographique de l'Ontario, conjointement avec la mise sur pied du système de références géographiques. La deuxième phase sera consacrée à l'automatisation de l'information toponymique actuellement contenue dans le système manuel. En prévision de cette automatisation, un sondage a été exécuté en 1984 pour déterminer les besoins des utilisateurs. Il ressort de cette étude que, de façon générale, les utilisateurs de données toponymiques s'intéressent en priorité aux éléments d'information suivants: les noms géographiques approuvés et enregistrés, l'étendue des éléments géographiques nommés ainsi que les lieux contenus dans les fichiers cartographiques de la province. À la lumière de ces résultats, on a recommandé une stratégie pour l'automatisation du système d'information sur les noms géographiques de l'Ontario. L'auteur décrit brièvement cette stratégie, de même que les arguments invoqués en sa faveur.

INTRODUCTION

On compte actuellement en Ontario près de 120,000 noms enregistrés d'éléments géographiques et de lieux habités. Ces noms sont conservés dans un système comprenant un fichier manuel et des cartes géographiques sur lesquelles sont portés les toponymes. Dès que les noms sont approuvés (près de la moitié le sont), ils font partie intégrante de la base de données cartographiques associée au système d'information géographique de la province. Ensemble, les fiches et les cartes forment l'unique répertoire complet de noms géographiques de l'Ontario et contiennent de l'information qui pourrait être utile à d'autres organismes.

Cependant, le système actuel est vieillot et difficile à utiliser. La recherche et le retrait de l'information prennent beaucoup de temps, et répondre aux demandes d'information toponymique est une tâche labo-

rieuse. Les fonds de carte sont plutôt fragiles, et il n'est pas facile d'en obtenir des exemplaires.

L'Ontario a donc décidé de modifier son système d'information toponymique, d'une part pour l'adapter au système d'information géographique de la province, et d'autre part pour accélérer le stockage et le retrait des noms enregistrés.

La première phase du programme de refonte du système est déjà en marche. Nous dressons actuellement une série d'index graphiques dans lesquels sont répertoriés tous les noms géographiques approuvés de l'Ontario, conjointement avec l'élaboration des cartes de base de l'Ontario (programme OBM) et avec l'établissement du système de références géographiques. L'information est portée sur des fonds de carte à 1/50 000 du Système national de référence cartographique (SNRC), qui couvrent une bonne partie de la province et qui contiennent déjà la plupart des toponymes. On a prévu d'établir les cartes en même temps que la production des cartes de base de l'Ontario, afin de produire les index graphiques sur pellicule stable dont on pourra ensuite tirer des diapositives (white prints) à peu de frais.

* Jeffrey R. Ball, spécialiste de la toponymie cartographique, Ministère des ressources naturelles de l'Ontario.

On distribue actuellement des exemplaires de la série aux bureaux régionaux et aux bureaux de district du ministère des Ressources naturelles, et on en met aussi à la disposition d'organismes provinciaux et fédéraux et d'autres utilisateurs intéressés.

La seconde phase du programme de refonte consistera à automatiser le système d'information sur les noms géographiques, pour accélérer l'accès à l'information actuellement contenue dans le système manuel et faciliter la révision de la base de données.

En prévision de cette seconde phase, la province a exécuté un sondage en 1984 pour connaître les besoins des utilisateurs. Je me propose ici de vous communiquer les résultats de ce sondage, après quoi je vous exposerai la stratégie qui a été recommandée pour l'automatisation du système d'information sur les noms géographiques de l'Ontario et les arguments invoqués en sa faveur.

MÉTHODE DE SONDEGE

Le questionnaire a été établi à la lumière de l'information recueillie au cours d'un sondage analogue fait en 1980. Celui-ci faisait partie d'une étude de faisabilité que nous avons commandée à une société d'experts-conseils pour savoir dans quelle mesure l'informatique pourrait nous aider à améliorer la diffusion des données toponymiques aux utilisateurs du système d'information de la province. Le deuxième sondage avait pour but de mettre à jour les renseignements issus de la première étude, avant d'entreprendre l'amélioration du système.

Très brièvement, le questionnaire en six parties dont nous nous sommes servis (voir l'appendice A) visait à établir:

- les besoins actuels et futurs en données toponymiques à caractère général et à caractère plus spécialisé;
- l'usage qui est fait de divers produits et services toponymiques, notamment les produits et services actuellement offerts par la province, et l'intérêt qu'ils suscitent.

En outre, nous avons invité les répondants à suggérer d'autres produits toponymiques qui seraient utiles selon eux. Dans une dernière question, nous avons tâché de connaître les organisations qui ont des bases de données automatisées contenant des noms géographiques.

Le questionnaire a été remis à vingt-et-un utilisateurs importants du système d'information toponymique de la province, que nous avons retenus d'après des données statistiques recueillies pendant et après l'étude de faisabilité. En tout, huit utilisateurs principaux et treize utilisateurs secondaires ont reçu un questionnaire.

Au nombre des utilisateurs principaux à qui un questionnaire a été envoyé, figuraient trois organismes toponymiques et cinq ministères provinciaux exerçant des activités en cartographie, en dessin ou en arpentage. Parmi les utilisateurs secondaires, on comptait un ministère provincial exécutant des travaux de cartographie, onze organisa-

tions participant à la gestion des ressources naturelles de l'Ontario et un ministère provincial chargé de diffuser de l'information générale sur le gouvernement.

Le taux de participation à ce sondage est renversant. Dix-sept questionnaires, soit 81 % du maximum possible, nous ont été renvoyés dûment remplis. Je profite d'ailleurs de cette occasion pour remercier tous les répondants.

RÉSULTATS DU SONDEGE

Besoins actuels et futurs en information

D'après les réponses données au questionnaire, il semblerait que les utilisateurs importants du système d'information toponymique de la province s'intéressent surtout

- aux noms enregistrés des lieux habités et aux noms approuvés des éléments géographiques (voir la figure 1);
- à la représentation graphique de l'étendue des éléments géographiques nommés et des lieux contenus dans les fichiers cartographiques de la province (voir la figure 2).

De plus, il semble assez évident, d'après les réponses, que les utilisateurs tirent généralement cette information des cartes, comme l'atteste la figure 3.

Le questionnaire nous a également fourni des renseignements utiles sur ce dont les utilisateurs auront besoin, dans l'avenir, en matière de noms géographiques. Les résultats donnent à penser que les organisations en quête d'information sur les noms géographiques pourraient accepter ces données sous d'autres formes que celles dans lesquelles elles leur sont actuellement livrées. Toutefois, ces résultats ne sont pas concluants, et cette question devra être réexaminée plus tard.

Autres constatations

Les figures 1 à 3 nous révèlent autre chose au sujet des besoins actuels en noms géographiques. Comme nous pouvons le voir dans l'histogramme de la figure

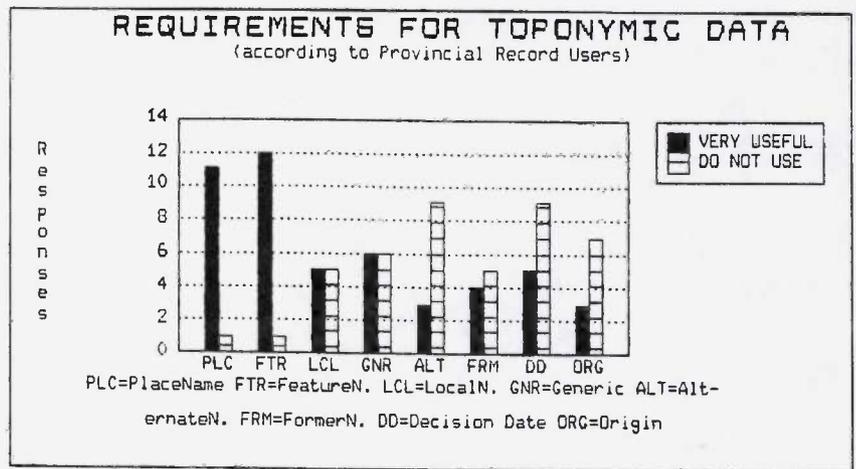


Figure 1

Degré d'utilisation des divers types de données toponymiques.

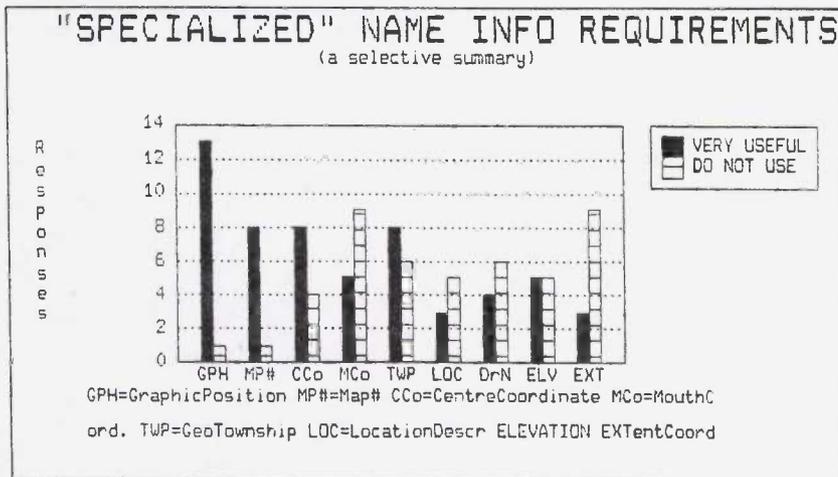


Figure 2 Degré d'utilisation de données toponymiques plus spécialisées.

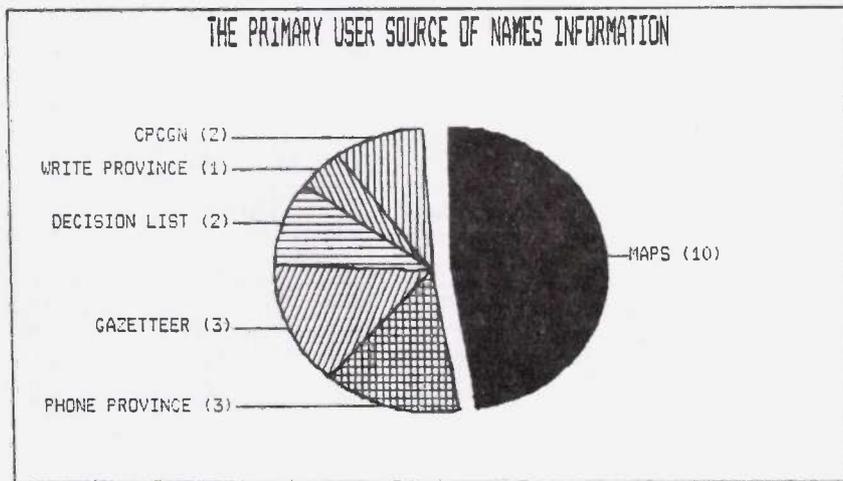


Figure 3 Principales sources d'information privilégiées par les utilisateurs. Le chiffre entre parenthèses correspond au nombre de répondants ayant désigné le produit ou le service indiqué comme la principale source d'information utilisée.

(à considérer comme le mode d'expression mathématique du lieu et de l'étendue des éléments géographiques), les descriptions de lieux et les descriptions de la direction de cours d'eau (deux méthodes de représentation qui emploient le langage courant) ont une utilité plutôt faible en comparaison des méthodes de représentation graphique des éléments géographiques.

STRATÉGIE RECOMMANDÉE POUR L'AUTOMATISATION DU SYSTÈME TOPONYMIQUE

Plusieurs stratégies peuvent être envisagées pour construire une base de données automatisée sur les noms géographiques. A la lumière des résultats du plus récent sondage, on a recommandé la stratégie suivante pour automatiser le système d'information sur les noms géographiques de l'Ontario:

a) La province de l'Ontario devrait procéder à l'automatisation de son système d'information sur les noms géographiques, en commençant par les cartes. Pour ce faire, elle pourrait introduire dans la base de données le nombre minimal de coordonnées nécessaires pour décrire avec précision l'étendue des éléments géographiques nommés et des lieux habités, tels qu'ils apparaissent sur la feuille de carte SNRC à 1/50 000, ou sur la feuille de carte OBM à 1/20 000 ou à 1/10 000, ou sur tous ces supports d'information.

b) S'il n'était pas possible d'automatiser tous les éléments de la partie graphique du système, il faudrait déterminer lesquels pourraient être codés le plus facilement; lesquels fourniraient de l'information sur l'emplacement et l'étendue des éléments géographiques désignés par des noms approuvés et des lieux habités désignés par des noms enregistrés (dont l'usage est approuvé); lesquels permettraient d'avoir accès rapidement à l'information présentement contenue dans le système manuel; lesquels faciliteraient la révision de la base de données; et lesquels seraient le plus compatibles avec la base de données du système de références géographiques de la province.

1, il ne semble pas y avoir encore une très grande demande de ce que nous pourrions appeler des "données toponymiques générales". En effet, les autres noms, les anciens noms, les noms locaux, les dates d'appellation officielle, les génériques et les origines sont, dans l'ensemble, peu demandés. Néanmoins, ce résultat peut porter à confusion. En fait, les utilisateurs importants de la base de données disent considérer les génériques et les dates d'appellation officielle comme d'une très grande utilité.

La demande de données plus spécialisées est illustrée dans la figure 2. On notera en particulier que quelques-unes des méthodes non graphiques les plus couramment utilisées pour véhiculer de l'information sur le lieu et l'étendue des éléments géographiques nommés ne sont pas très en demande. De plus, les coordonnées

c) Outre l'information portée sur les cartes, la version originale de la base de données automatisée devrait contenir tous les noms approuvés d'éléments géographiques et tous les noms enregistrés de lieux habités actuellement contenus dans la base de données de la province.

d) Finalement, la province devrait considérer la possibilité d'introduire dans la base de données les génériques, les dates d'appellation officielle et les désignations des townships pour tous les toponymes approuvés et enregistrés dont il est question ci-dessus.

RAISONS INVOQUÉES POUR CENTRER LE PROCESSUS D'AUTOMATISATION SUR LES CARTES

Voici quelques-unes des raisons les plus importantes qui ont été invoquées pour que la base de données automatisée soit constituée à partir des cartes du système actuel:

- a) Les utilisateurs importants de la base de données seront nettement avantagés par cette approche de l'automatisation, puisque les cartes du système d'information toponymique actuel de la province contiennent les renseignements les plus en demande, c'est-à-dire les noms approuvés et enregistrés, ainsi que la description et l'emplacement des éléments géographiques et des lieux désignés par ces noms.
- b) C'est des cartes que les utilisateurs tirent la majeure partie de l'information présentement contenue dans les fiches, dans le répertoire géographique de l'Ontario et dans les listes de décisions de la Commission des noms géographiques de l'Ontario. Les cartes se sont avérées jusqu'ici, semble-t-il, des supports d'information utiles et durables.

c) Non seulement une base de données toponymiques fondée sur des cartes pourra produire de l'information sous forme graphique, mais on pourra en tirer également des produits et des données non graphiques, par exemple un répertoire des noms géographique de l'Ontario, ainsi que des descriptions de lieux et des renseignements sur la direction de cours d'eau.

d) Il est probable qu'une base de données toponymiques contenant suffisamment de données qui décrivent sous forme graphique des éléments naturels et des lieux habités sera davantage compatible avec les puissantes bases de données cartographiques numériques actuellement en construction que ne le sont les autres systèmes privés de cette information.

En résumé, non seulement une base de données toponymiques construite autour d'un noyau cartographique caractérisé par une grande stabilité (et que je considère de surcroît comme la source traditionnelle d'information sur les noms géographiques) pourrait satisfaire les besoins actuels des utilisateurs, mais elle serait aussi suffisamment souple pour être augmentée en fonction des besoins des utilisateurs.

APPENDICE A: QUESTIONNAIRE SUR L'UTILISATION DES NOMS GÉOGRAPHIQUES

Veuillez trouver ci-joints:

- (1) Un questionnaire sur l'utilisation des noms géographiques.
- (2) Un spécimen d'une liste de décisions de la Commission des noms géographiques de l'Ontario. (Voir le questionnaire, questions 1.1, 1.2 et 2.3.)

(3) Un spécimen d'un "Index graphique", indiquant les noms géographiques approuvés de l'Ontario. (Voir le questionnaire, questions 1.1, 1.2 et 2.3.)

Nous vous prions de bien vouloir remplir le questionnaire et de nous le retourner dans l'enveloppe déjà adressée et affranchie que vous trouverez sous pli.

QUESTIONNAIRE - UTILISATION DES NOMS GÉOGRAPHIQUES

Nom :
Organisation :
Service :
Adresse :
N° de tél. :

Nous vous demandons de répondre aux questions suivantes en donnant le point de vue de votre organisation, plutôt que le vôtre. En d'autres termes, vous devez nous dire comment, à votre connaissance, votre organisation utilise les noms géographiques.

1.1 UTILISATION ACTUELLE DU SYSTÈME D'INFORMATION TOPONYMIQUE PROVINCIAL

Veuillez cocher la case appropriée.

Question	Oui	Non	Ne sais pas
(i) Utilisez-vous les listes de décisions de la Commission des noms géographiques de l'Ontario (CNGO)?			
(ii) Utilisez-vous les index graphiques?			
(iii) Téléphonnez-vous à la Section des noms géographiques (SNG) pour obtenir de l'information?			
(iv) Écrivez-vous à la Section des noms géographiques pour obtenir de l'information?			

1.2 Veuillez encercler les produits ou services que vous aimeriez utiliser ou continuer à utiliser.

Listes de décisions
de la CNGO

Index
graphiques

Contacts par
téléphone avec
la SNG

Contacts par
écrit avec
la SNG

2.1 UTILISATION ACTUELLE D'AUTRES SOURCES D'INFORMATION TOPONYMIQUE

Veuillez cocher la case appropriée. Vous pouvez ajouter d'autres sources dans les espaces prévus à cette fin.

Actuellement, votre organisation tire ses renseignements sur les noms géographiques:	Oui	Non	Ne sais pas
(i) de cartes?			
(ii) du Répertoire?			
(iii) de l'organisme fédéral chargé des noms géographiques?			
(iv) de _____ !			
(v) de _____ !			
(vi) de _____ !			

2.2 Veuillez indiquer, par ordre de préférences, lesquels parmi les produits ou services mentionnés dans les questions 1 et 2 sont les plus utiles à votre organisation.

3.1 BESOINS ACTUELS ET FUTURS EN DONNÉES TOPONYMIQUES DE BASE

Veuillez cocher la case appropriée. Il s'agit, dans un premier temps, de déterminer vos besoins actuels en données toponymiques. Il sera question de vos besoins futurs dans la question 3.2.

Genres de données requis	Très Utile	Utile	Utilisé occasionnellement	Inutilisé
(i) Noms de lieux				
(ii) Noms d'éléments géographiques				
(iii) Autres noms (par ex. dans une autre langue)				
(iv) Anciens noms				
(v) Noms d'usage local				
(vi) Dates d'appellation officielle				
(vii) Génériques (dans les cas où le générique n'est pas évident, comme dans Red Lake (lieu))				
(viii) Origine (par ex. l'histoire du nom)				

3.2 Parmi les genres de données que, dans la réponse à la question précédente, vous qualifiez de "inutilisés" ou "utilisés occasionnellement", lequel ou lesquels pourraient être utiles à votre organisation. Plutôt que de les écrire au long, vous pouvez n'indiquer que les chiffres romains correspondants.

4.1 BESOINS ACTUELS ET FUTURS EN DONNÉES SPÉCIALISÉES

Veillez cocher la case appropriée. Vous pouvez vous servir du verso de cette feuille (et de la suivante) pour préciser vos réponses au besoin ou formuler d'autres observations; veuillez dans ce cas utiliser les chiffres romains pour bien indiquer de quels genres de données il s'agit. Dans un premier temps, nous nous intéresserons à vos besoins actuels en données toponymiques. Vos besoins futurs feront l'objet de la question suivante.

Genres de données requis	Très Utile	Utile	Utilisé occasionnellement	Inutilisé
(i) Noms et représentation graphique de son application (par ex., carte indiquant les noms et l'étendue des éléments géographiques nommés).				
(ii) Numéro de la carte sur laquelle figure un élément nommé. [Prière d'indiquer la série].				
(iii) Plus d'un numéro de carte (par ex. si l'élément s'étend sur plus d'une feuille de carte de la même série).				
(iv) Coordonnées du centre des éléments/lieux nommés. [Prière de préciser le système de coordonnées requis (par ex. coordonnées géographiques, coordonnées UTM, etc) et le degré de précision (par ex., au degré, à la minute ou à la seconde le plus près; à 10 000 m, 1 000 m ou 100 m près).]				

Genres de données requis	Très Utile	Utile	Utilisé occasionnellement	Inutilisé
(v) Coordonnées de l' <u>embouchure</u> , dans le cas d'un cours d'eau.				
(vi) Coordonnées de la <u>source</u> , dans le cas d'un cours d'eau.				
(vii) Coordonnées <u>limites</u> (étendue) d'éléments/de lieux.				
(viii) Désignation <u>géographique</u> du township dans lequel l'élément/le lieu a son centre.				
(ix) Plusieurs désignations géographiques de townships (par ex. dans le cas où l'élément s'étend sur plus d'un township).				
(x) Désignation du comté, du district, de la municipalité régionale ou de la municipalité de district dans lequel l'élément/le lieu a son centre.				
(xi) Plusieurs désignations de comtés, de districts, de municipalités régionales ou de municipalités de district.				
(xii) <u>Localisation</u> , telle que donnée dans le Répertoire géographique (par ex. Cronin Lake: NW de Trout Lake).				
(xiii) Description de l'emplacement par des distances (par ex. Cronin Lake: 1,6 km/1 mille au NW de Trout Lake).				
(xiv) Direction des cours d'eau (par ex. ruisseau Little: coule vers le NW et se jette dans le lac Big).				
(xv) Altitude des éléments/lieux nommés.				

4.2 Parmi les genres d'information que vous avez qualifiés de "inutilisés" ou "utilisés occasionnellement", lequel ou lesquels, selon vous, pourraient être utiles à votre organisation. (Nous vous demandons une fois de plus d'écrire les chiffres romains plutôt que les genres d'information au long).

5.1 INTÉRÊT POUR DES PRODUITS ET SERVICES TOPONYMIQUES POSSIBLES

Dans la question ci-dessous, nous tentons de déterminer dans quelle mesure votre organisation est intéressée par des produits et services toponymiques POSSIBLES. Nous insistons sur ce dernier mot, parce qu'il faut bien comprendre que les produits énumérés ci-dessous ne sont pas encore offerts par l'organisme toponymique provincial et qu'ils ne le seront pas nécessairement. Les réponses à cette question nous permettront néanmoins de mettre sur pied un service efficace d'information sur les noms géographiques.

5.2 Veuillez cocher la case appropriée:

Produit possible	Serait très utile	Serait Utile	Serait utile à l'occasion	N'aurait aucune utilité
(i) Listes alphabétiques de noms géographiques avec coordonnées du centre/de l'embouchure (seulement), classées selon la région de votre choix. Nota: Ces listes ne comprendraient que les éléments géographiques et les lieux nommés dont le centre (pour les lacs) et l'embouchure (pour les cours d'eau) seraient situés dans la région choisie.				
(ii) Listes alphabétiques de noms géographiques avec plusieurs coordonnées pour définir l'étendue de l'élément (au besoin), classées selon la région* (approximative) de votre choix. Nota: Elles renfermeraient essentiellement tous les éléments géographiques et lieux (au besoin) qui se trouvent dans la région indiquée, en plus de quelques éléments et lieux se trouvant au voisinage immédiat de la région.				
(iii) Listes de noms semblables à (i) et (ii), mais classées par génériques ou par dates d'appellation officielle, ou selon les deux critères à la fois.				
(iv) Listes ci-dessus, mais sur <u>microfiches</u> .				
(v) Cartes-transparents montrant les noms, les génériques et les coordonnées limites de l'élément géographique; région et échelle au choix de l'utilisateur.				
(vi) Information toponymique sous forme lisible par ordinateur.				

* La région utilisée serait définie d'après les feuilles de carte de base de l'ensemble de l'Ontario et serait donc un peu plus grande que la région choisie par l'utilisateur.

5.3 Veuillez mentionner ci-dessous tout autre produit ou service toponymique qui serait très utile à votre organisation.

6.1 Votre organisation dispose-t-elle déjà d'une base de données automatisée contenant de l'information toponymique? Veuillez encrer la bonne réponse.

Oui Non Je ne sais pas

MERCI DE VOTRE PARTICIPATION.

VERS UN NOUVEAU SYSTÈME

Commission de toponymie du Québec*

RÉSUMÉ. La Commission de toponymie ayant été dans l'impossibilité de se doter du système automatisé de gestion de données toponymiques (ONOMA) à cause des coûts prohibitifs principalement, elle a dû opter pour l'amélioration du système informatique qu'elle utilisait depuis une douzaine d'années. Ce système amélioré fut baptisé STRATO et il signifie "système de traitement des données toponymiques officielles". Les principales fonctions de ce système sont intitulées: correction, sélection, statistiques, gazette officielle, tables, le fichier des toponymes populaires et le système de gestion documentaire. Pour répondre à ses besoins dans un contexte de gestion de l'information et d'accroissement des services à la clientèle, la Commission sera dotée de micro-ordinateurs dès 1986. Ce transfert technologique vers la micro-informatique apportera à la Commission l'autonomie d'opération et de gestion d'accessibilité presque immédiate de l'information.

RAPPEL HISTORIQUE

Lors du Premier colloque sur l'automatisation des noms géographiques tenu à Ottawa, en 1980, la Commission de toponymie avait présenté une communication intitulée SYSTÈME AUTOMATISÉ DE GESTION DE DONNÉES TOPONYMIQUES.¹

Le représentant de la Commission, Pierre Barabé, avait décrit dans le détail les caractéristiques du système ONOMA. Tel que nous l'avions précisé à l'époque, il s'agissait d'un projet dont la réalisation pourrait être sensiblement différente de ce que l'on projetait. Le temps devait donner raison à cette prédiction. En effet, même si la Commission de toponymie avait obtenu un accord de principe du Conseil du trésor québécois sur la structure générale du système ONOMA, la Commission devait poursuivre ses études sur les mécanismes qui devaient être mis sur pied pour permettre l'intégration des différents corpus toponymiques existants, sur les ressources requises pour effectuer l'intégration des données contenues dans la propre banque de la Commission et surtout, elle devait procéder à une étude coûts/bénéfices quant au choix de logiciel à utiliser.

Les mois qui suivirent le Premier colloque devaient s'avérer déterminants pour l'orientation que la Commission pouvait escompter donner à son projet ONOMA. Il lui fallut se rendre à l'évidence qu'elle n'avait tout simplement pas les moyens, plus particulièrement les ressources monétaires suffisantes, pour se doter du système ONOMA. Elle fit appel à des consultants extérieurs pour se voir confirmer que, effectivement, l'investissement nécessaire pour le démarrage, le développement, l'entretien et l'exploitation du système ONOMA étaient prohibitifs particulièrement en période de gestion de la décroissance.

BESOINS TOPONYMIQUES

La Commission de toponymie devait toujours se préoccuper de l'administration de l'ensemble des noms de lieux du territoire québécois, d'officialiser les toponymes, de les diffuser... Elle se devait d'appliquer ses juridictions en matière de critères de choix des toponymes, des règles d'écriture de ceux-ci, de terminologie géographique... loin de diminuer, les besoins de supports techniques de la Commission progressaient rapidement. A titre indicatif, mentionnons que le 30 avril 1985, elle possédait 122 133 enregistrements toponymiques dans son fichier maître et que chaque enregistrement était susceptible de générer un maximum de 23 informations (Annexe 1), une amélioration approximative de 30 % au cours des cinq dernières années. Le fichier maître qui est en croissance constante, gère donc présentement un total de 2 578 599 informations.

Ainsi, ayant dû surseoir au projet ONOMA, elle opta pour l'amélioration du système informatique qui, depuis une douzaine d'années, lui permettait notamment de produire la Gazette officielle du Québec et certaines listes de toponymes officiels regroupés selon différents paramètres dont, notamment, l'ENTITÉ, le NOM, la LOCALISATION, le numéro de CARTE dans le système de référence cartographique national (Annexe 2). Cette deuxième génération du système informatique fut baptisé STRATO qui signifie "système de traitement des données toponymiques officielles". Pour améliorer les performances et répondre au volume croissant de ses besoins informatiques du traitement des toponymes, la Commission a créé une nouvelle unité administrative le SERVICE DE DÉVELOPPEMENT DES SYSTÈMES. Aux besoins quotidiens suscités par l'administration courante de la toponymie, étaient venus se greffer l'exploitation d'un fichier spécialisé sur un type particulier de toponymes soit les TOPONYMES POPULAIRES et certains besoins de GESTION DOCUMENTAIRE.

SERVICE DE DÉVELOPPEMENT DES SYSTÈMES

Le Service de développement des systèmes a comme activité prioritaire l'entretien et l'exploitation des systèmes informatiques actuels de la Commission de toponymie soit le système de traitement de données toponymiques officielles (STRATO), le fichier des toponymes

* Présenté par Jean Poirier, adjoint au président, Commission de toponymie du Québec.

1 Voir CANOMA, vol. 6, n° 1, juillet 1980.

populaires (FTP) ainsi que le système de gestion documentaire (SGD).

STRATO

Communément appelé fichier des toponymes, ce système reflète en quelque sorte le **Répertoire toponymique officiel du Québec**. Il contient donc tous les toponymes parus à la **Gazette officielle du Québec** et, au 30 avril 1985, il contenait 122 133 toponymes et odonymes (noms de voies de communication). Si ce système a pour but principal la production annuelle de la **Gazette officielle du Québec**, de nombreux autres rapports tels des rapports statistiques et des listes sélectives sont aussi produits.

Tous les programmes et procédures se regroupent dans les six (6) principales fonctions suivantes:

- **SUPPLÉMENT**

La fonction SUPPLÉMENT a pour rôle d'ajouter des enregistrements au fichier maître. Cette opération s'effectue en deux temps: le contenu des fiches est d'abord saisi au terminal pour constituer un fichier de saisie, et le contenu de ce dernier est ensuite ajouté au fichier maître.

- **CORRECTION**

La fonction CORRECTION sert à apporter des modifications à des enregistrements du fichier maître. Cette opération se fait aussi en deux temps: les corrections sont d'abord saisies au terminal, formant un fichier de transactions et ces transactions sont ensuite traitées par le programme de mise à jour qui fait les corrections au fichier maître.

- **SÉLECTION**

La fonction SÉLECTION extrait des enregistrements du fichier maître pour produire des listes de toponymes, des fichiers extraits ou simplement pour les compter. Les critères de sélection ne sont limités que par le nombre de champs dans le fichier.

- **STATISTIQUES**

La fonction STATISTIQUES produit des statistiques sur le contenu du fichier maître ou de tout autre fichier extrait du fichier maître. Les statistiques produites ne sont que des compteurs cumulés selon différentes catégories d'information.

- **GAZETTE OFFICIELLE**

La fonction GAZETTE OFFICIELLE sert à produire un fichier utilisable par une maison de photocomposition. Ce fichier contient l'information qui doit être publiée chaque année à la **Gazette officielle du Québec** (i.e. les toponymes officialisés depuis la dernière parution et toutes modifications apportées aux toponymes officialisés antérieurement).

- **TABLES**

La fonction TABLES regroupe toutes les opérations

relatives aux tables de codification: création, mise à jour et liste. Le système comprend six tables: cantons, divisions de recensements, entités, codes géographiques, municipalités régionales de comté (MRC) et régions administratives.

FTP

Le fichier des toponymes populaires a pour but de répondre aux demandes d'information concernant les toponymes en rapport avec les divisions territoriales du Québec.

Pour la période relative à l'exercice financier 1984-1985, plus de 80 rapports ont été distribués aux usagers de ce service de la Commission de toponymie.

SGD

Le système de gestion documentaire implanté pendant l'année 1983-1984 a pour but de veiller à la sauvegarde et à la mise à jour des fichiers concernant les données descriptives des documents utilisés par les différentes unités administratives de la Commission.

Pour la période relative à l'exercice financier 1984-1985, plus de 1 300 transactions ont été effectuées.

MODIFICATIONS AU SYSTÈME

Certaines modifications ont été apportées au système informatique actuel afin de permettre à la Commission de compléter l'information déjà disponible relativement à la toponymie officielle du Québec.

Une des principales améliorations apportées au système a été la création et la mise à jour des tables des codes géographiques (municipalités), des MRC (municipalités régionales de comté) et des RA (régions administratives).

Au fichier maître, quelques ajustements à des fonctions existantes sont venues perfectionner l'état existant de ce fichier. Mentionnons l'ajout d'une deuxième paire de coordonnées pour une plus grande précision, l'ajout de la classe d'entité qui permet maintenant de distinguer entre l'enregistrement d'un odonyme et d'un toponyme, ainsi que des modifications mineures à la division de recensement et aux cantons.

MICRO-INFORMATIQUE

La Commission de toponymie est dans un processus de réflexion et de recherche au niveau informatique dans le but de trouver une solution technologique et organisationnelle globale permettant de répondre à ses besoins dans un contexte de gestion de l'information et d'accroissement des services à la population.

L'implantation de la micro-informatique impliquera premièrement le transfert de toute l'information contenue dans le système actuel (STRATO) sur micro-ordinateur avec tout ce que cela comporte d'installation physique, le logiciel, de formation du personnel ainsi que la mise sur pied d'un système de sécurité.

L'utilisation en parallèle des deux systèmes au début du processus de changement permettra à la Commission de poursuivre ses activités normales (saisie

et mise à jour sur l'ancien système) tout en se laissant une marge de sécurité pendant la période d'adaptation au nouveau système.

Deuxièmement, une confrontation des deux fichiers de données sera faite ultérieurement et les ajustements nécessaires y seront apportés pour finalement abandonner l'ancien système dès que le nouveau sera rôdé.

Cette recherche d'autonomie d'opération et de gestion d'accessibilité presque immédiate de l'information, la continuité dans l'élaboration de projets, la possibi-

lité d'utilisation de logiciels adaptés aux besoins de la Commission ainsi que la diminution des coûts d'exploitation à court et à long terme, ne sont qu'un aperçu des quelques avantages du point de vue technologique et organisationnel que ce transfert technologique vers la micro-informatique apportera à la Commission de toponymie.

Nous espérons que cette approche rationnelle puisse faire partie du vécu quotidien de la Commission de toponymie d'ici un an.

ANNEXE 1

TYPE DE TRANSACTION EXISTANT DANS LE SYSTÈME

<u>Transaction</u>	<u>Avant 1985</u>	<u>1985</u>	<u>Transaction</u>	<u>Avant 1985</u>	<u>1985</u>
Numéro séquentiel	1	1	Classe d'entité (odonyme/toponyme)	0	2
Coordonnées	1 paire	2 paires	Division de recensement	1	4
Codification linguistique	1	1	Code géographique	0	4
Code sémantique	1	1	Canton	3	4
Date de publication	1	1	Toponyme	1	1
Code d'entité	1	1	Feuille	1	1
			Renvoi	1	1

ANNEXE 2

AJOUT DE NOUVEAUX TOPONYMES AU FICHIER

- Information à saisir (au minimum)
 - toponyme
 - numéro de feuille
 - code d'entité
 - classe d'entité (type: odonyme/toponyme)
 - coordonnée (une paire)
 - code géographique (minimum 1; maximum 4)
- division de recensement (minimum 1; maximum 4)
- un renvoi
- Informations supplémentaires
 - canton
 - date de publication
 - coordonnées (2e paire)
 - codification linguistique
 - codification sémantique

SITUATION AUTOMATIQUE D'UN TOPONYME DANS SA MUNICIPALITÉ

Clément Nolette*

RÉSUMÉ. Ce texte présente la démarche suivie pour situer automatiquement les toponymes du Québec dans leur municipalité respective à partir des coordonnées géographiques des toponymes et d'un fichier numérique des limites municipales.

A la demande de la Commission de toponymie du Québec, nous avons développé un programme permettant d'ajouter de façon automatique le code de municipalité aux toponymes contenus dans leur banque. D'une part les coordonnées géographiques des points d'implantation des toponymes faisaient déjà partie de cette banque et d'autre part nous savions que le fond de carte numérisé des limites municipales du Québec était disponible depuis peu au Ministère de l'Énergie et des Ressources. Tous les éléments requis pour réaliser ce projet étaient donc disponibles, il suffisait maintenant de créer le lien entre ces deux fichiers. Un souci d'efficacité nous a amené à découper le travail en deux étapes, soit une étape de préparation de fichiers modifiés et une étape de traitement. A l'intérieur de celle-ci, un second souci, celui de l'exactitude, nous a conduit à créer un mécanisme permettant de classer les résultats obtenus en trois fichiers distincts selon leur fiabilité. Une brève description des algorithmes utilisés suit la présentation du système.

Lors de la première étape, nous sommes partis d'une version du fichier des toponymes d'où avaient été exclus les odonymes et les hydronymes, étant donné le caractère linéaire de ce qu'ils nomment, ainsi que les toponymes dont le code de municipalité était déjà présent. Nous avons extrait de ce fichier uniquement l'information nécessaire pour nos besoins afin de réduire la taille du fichier traité. Nous avons donc conservé, pour chaque toponyme, son numéro de référence, ses coordonnées, le code de sa division de recensement, le numéro de son feuillet cartographique et les quinze premiers caractères de son nom. Afin de faciliter le traitement mathématique ultérieur, nous avons toutefois modifié la forme des coordonnées et utilisé des degrés avec partie fractionnaire au lieu des degrés avec minutes et secondes pour représenter les latitudes et les longitudes. Le numéro du feuillet cartographique et les premiers caractères du nom ont été retenus afin de faciliter la tâche de la personne qui aura à vérifier les cas difficiles que le programme n'aura pu classer de façon sûre.

Le fichier résultant de cette extraction est ensuite trié en ordre croissant des numéros de division de recensement. Tous les toponymes se trouvent ainsi regroupés en ensembles consécutifs correspondant à chacune des divisions (voir Figure 1). Le même tri sera par la suite appliqué au fichier des municipalités de sorte qu'il sera possible de traiter les toponymes division par division. Tout ceci vise à réduire le temps de re-

cherche en limitant le nombre de municipalités pour lesquelles le programme devra vérifier si le toponyme se trouve à l'intérieur ou non.

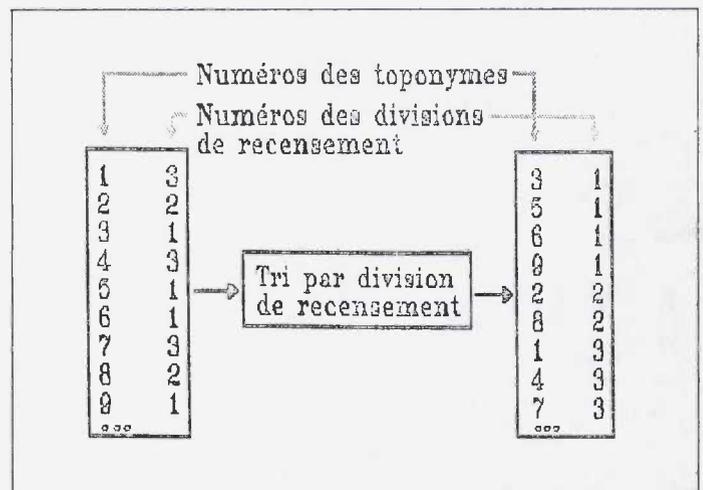


Figure 1

Le fichier des municipalités est un fichier F.I.L.A.M. (Fichier Informatisé des Limites Administratives Municipales) fourni par le Ministère de l'Énergie et des Ressources du Québec. Ce fichier contient toutes les limites municipales de la province, et chaque municipalité est représentée par un polygone auquel est rattaché un code géographique. Il s'agit précisément du code que nous voulons ajouter aux dossiers des toponymes situés dans cette municipalité, dans ce polygone. Dans le cas de ce fichier aussi, nous avons transformé les coordonnées géographiques, latitudes et longitudes, en degrés avec partie décimale au lieu de degrés avec minutes et secondes. D'ailleurs, par la suite nous utilisons directement ces coordonnées de type sphérique et les assimilons à des coordonnées cartésiennes étant donné que la relation d'inclusion ou d'exclusion d'un point par rapport à un polygone n'aurait pas été modifiée par l'algorithme de projection que l'on aurait pu appliquer aux coordonnées des toponymes et à celles des municipalités.

Avant de procéder au tri du fichier des municipalités par division de recensement, nous avons effectué certains calculs nous donnant, pour chaque municipalité, les limites du carré circonscrit à son polygone et la surface carré. Ces nouvelles données ont été ajoutées

* Clément Nolette, étudiant gradué, département de géographie, Université Laval.

au fichier des municipalités et lors du tri par division, un second tri est effectué, à l'intérieur de chaque division, en ordre croissant selon la surface des carrés circonscrits. Le besoin de ce second tri s'explique par la façon dont la recherche s'effectue pour localiser un toponyme dans une des municipalités de sa division de recensement. En effet, cette recherche est séquentielle et, pour éviter que des toponymes situés dans une municipalité imbriquée dans une autre plus vaste ne soient situés par erreur dans cette dernière, nous devons nous assurer que la vérification d'inclusion se fasse dans un ordre où les municipalités de petite surface seront traitées avant celles de grande surface (voir Figure 2).

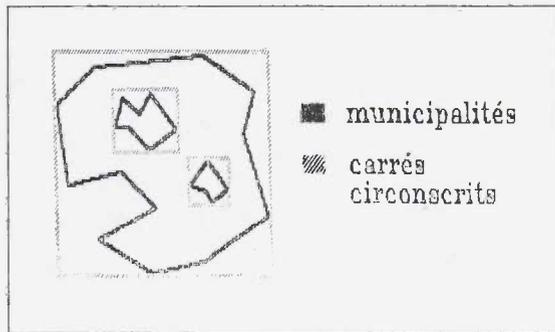


Figure 2

Une fois cette étape de préparation du fichier des toponymes et du fichier des municipalités complétée, nous passons au traitement proprement dit (voir Figure 3). Celui-ci produit trois fichiers, tous contiennent l'information retenue lors de l'extraction effectuée dans la première étape, plus le code de municipalité approprié si la recherche s'est bien déroulée. Le premier fichier contient les toponymes sûrs, c'est à dire ceux dont les points d'implantation ont été situés de façon certaine dans une municipalité. Le deuxième contient les toponymes moins sûrs, soit ceux dont les points d'implantation sont situés en deçà d'une distance critique

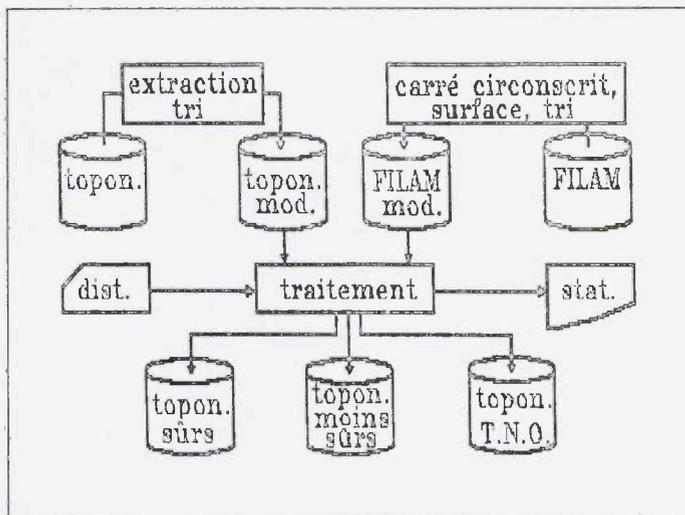


Figure 3

des limites des municipalités trouvées pour ces toponymes; cette distance est paramétrique et peut être modifiée au besoin. Une vérification manuelle devrait normalement être faite pour les toponymes placés dans ce fichier. Enfin, le troisième fichier contient les toponymes n'ayant pu être situés dans aucune municipalité, c'est à dire ceux appartenant aux T.N.O. (Territoires Non Organisés). Certaines informations statistiques sont aussi fournies par le programme de traitement, dont le nombre de toponymes sûrs et moins sûrs par municipalité ainsi que le nombre de toponymes dans les T.N.O. par division de recensement.

Voici maintenant une brève description du mode de fonctionnement du programme de traitement. Celui-ci amène d'abord en mémoire toute l'information rattachée aux municipalités d'une division de recensement: codes géographiques, carrés circonscrits, surfaces et coordonnées. Il lit ensuite les toponymes un à un et les traite au fur et à mesure. Lorsqu'il rencontre un toponyme dont le code de division de recensement est différent de celui des municipalités, il effectue alors la lecture des informations rattachées aux municipalités de la prochaine division, puis reprend le traitement des toponymes; ainsi de suite, jusqu'à la fin du fichier des municipalités et de celui des toponymes.

Le traitement d'un toponyme consiste à vérifier si son point d'implantation se situe dans une des municipalités de sa division de recensement. Cette vérification se fait pour chaque municipalité jusqu'à ce que l'on trouve celle où il se situe. Si aucune municipalité n'est trouvée, le point appartient à un T.N.O. Un premier test permet à vérifier rapidement si un point est exclu d'une municipalité, il suffit de vérifier s'il est situé à l'intérieur ou à l'extérieur de son carré circonscrit (voir Figure 4). S'il est situé à l'extérieur, il ne peut évidemment pas être dans cette municipalité et on passe à la municipalité suivante. Par contre, s'il est

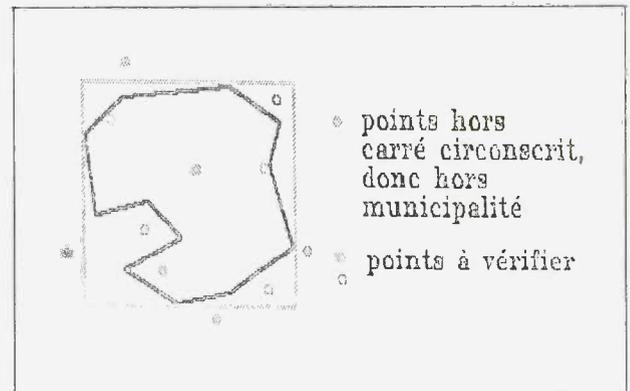


Figure 4

à l'intérieur, on doit procéder à un test plus complexe d'inclusion d'un point dans un polygone, celui-ci est décrit au prochain paragraphe. Si le point n'est pas dans le polygone, on passe à la municipalité suivante. Sinon on calcule la distance du point aux différents segments constituant la limite de la municipalité et la distance aux sommets constitués par la rencontre de ces segments dans le cas des angles de plus de 180 degrés. Si on rencontre une distance inférieure à la distance critique fournie comme paramètre, le toponyme sera classé comme moins sûr, sinon il sera considéré comme sûr (voir Figure 5).

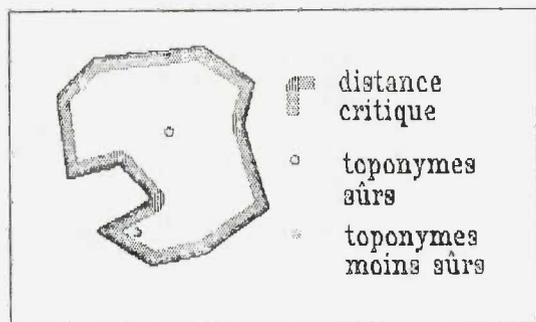


Figure 5

L'algorithme de vérification d'inclusion d'un point dans un polygone est basé sur le principe qu'une ligne droite traverse toujours aucun, ou un nombre pair de côtés d'un polygone situé dans un même plan, sauf si elle passe par un sommet. Nous prenons donc une ligne verticale passant par le point dont nous voulons tester l'inclusion et nous vérifions d'abord si elle passe aussi par un des sommets du polygone, dans ce cas, un léger décalage de l'abscisse du point nous permet de continuer le test après reprise de cette vérification. Nous calculons ensuite les points d'intersection entre cette droite verticale et les différents segments du polygone qu'elle croise. Enfin nous comptons le nombre de points d'intersection situés au dessus ou au dessous du point. Si le nombre est impair, le point est à l'intérieur du

polygone, sinon, si le nombre est pair, le point est à l'extérieur (voir Figure 6).

Au moment de la rédaction de ce texte, le système a été testé avec 23 municipalités de deux divisions de recensement de la région de Charlevoix et tous les toponymes de cette région. Tous les toponymes déclarés sûrs le sont effectivement, mais plus de 80 % des moins sûrs sont bien localisés. Il s'agit donc de calibrer la distance critique afin que ces derniers deviennent sûrs, tout en s'assurant qu'aucun cas douteux ne s'y mêle. Quelques toponymes classés dans des T.N.O. n'ont toutefois pas raison d'y être et nous sommes présentement à la recherche de la solution à ce problème.

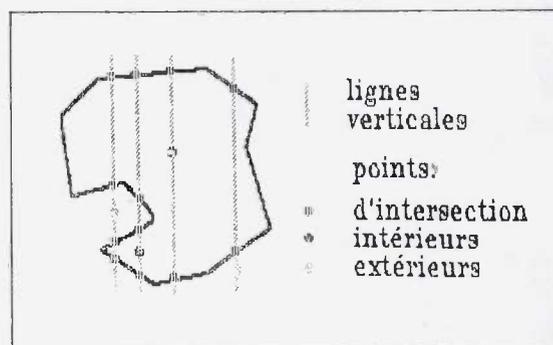


Figure 6

"GEOGRAPHIC NAMES INFORMATION SYSTEM (GNIS)"

PRINCIPES ET FONCTIONS

Roger L. Payne*

[Translation]

RÉSUMÉ. Le système GNIS "Geographic Names Information System", (ou système d'information sur les noms géographiques) est le fruit d'un long travail de saisie et de publication de noms géographiques. Il s'agit essentiellement d'un système toponymique automatisé et polyvalent, capable de satisfaire une foule de besoins à tous les niveaux de l'administration et dans le secteur privé. Conçu par des toponymistes, il sert en quelque sorte de répertoire des noms géographiques des États-Unis; de ce fait, il apporte une aide précieuse au National Mapping Program de l'U.S. Geological Survey (USGS) et constitue désormais un outil fondamental des recherches et des applications toponymiques.

INTRODUCTION

Le recensement des noms des entités géographiques sous forme de répertoires et d'index géographiques constitue, depuis le 18e siècle, un moyen pratique de repérer des lieux, des accidents géographiques et des

* Roger L. Payne, Chef, Geographical Names Information Section, National Mapping Division, United States Geological Survey, Reston, Virginia.

régions aux États-Unis. Or, pendant la grande période de colonisation des 18^e et 19^e siècles, le nombre de toponymes s'est accru au point qu'il est vite devenu impossible de tenir à jour les répertoires, même à l'échelle régionale ou locale. Les États et l'administration fédérale ont fait paraître beaucoup de répertoires au 19^e siècle, mais ceux-ci contenaient surtout les noms des lieux habités et des entités géographiques d'importance. L'exhaustivité était généralement tenue pour un objectif difficile voire impossible à atteindre. En 1892, l'USGS commença à cataloguer des noms géographiques dans le cadre du programme cartographique national; c'est ainsi que les répertoires de 12 États et de 3 territoires furent publiés entre 1894 et 1906. Malgré le souci d'exactitude et d'exhaustivité, ces répertoires ne renfermaient que les noms figurant sur les cartes topographiques de l'USGS, dont la plupart étaient établies aux échelles de 1/125 000 (surface de 30 minutes) et de 1/63 360 (surface de 15 minutes). Beaucoup de noms locaux désignant des entités de moindre importance étaient omis. En outre, au tournant du siècle, un petit pourcentage seulement de la superficie des États-Unis avait été porté sur des cartes. Faute de cartes, on se résigna donc à abandonner le programme.

ÉTABLISSEMENT D'UN RÉPERTOIRE NUMÉRIQUE

Le principe de l'établissement d'un système toponymique automatisé ne date pas d'aujourd'hui, et on n'est pas prêt de l'abandonner. Pendant la première moitié du 20^e siècle, le nombre de cartes topographiques à grande échelle des États-Unis s'est accru considérablement, mais pas au point de rendre possible la saisie systématique des noms géographiques pour l'établissement de répertoires toponymiques. Il y avait certes des travaux qui se faisaient à l'échelle régionale ou locale, mais rien n'était organisé à la grandeur du pays et, souvent, le travail était confié à des gens qui n'avaient pas la formation requise. Pour être utile, un répertoire ne devait pas se limiter à indiquer le nom de l'entité, ainsi que l'État et le comté dans lesquels il se trouvait. On devait aussi inscrire la position de l'entité, exprimée dans un système de quadrillage avec latitude et la longitude exactes à quelques secondes près. De plus, pour déterminer rapidement la position d'une entité, il fallait un système de référence qui permettrait de retrouver la feuille de carte sur laquelle l'entité était représentée. Des compléments d'information, comme les variantes et les formes secondaires des noms, faciliteraient les références et applications historiques et donneraient aussi une idée de l'évolution du toponyme. Dans l'établissement des premiers répertoires, on ne se souciait guère de l'exactitude de la position, jugée superflue compte tenu de l'usage qui était fait du produit. Pour être complète, la liste de noms devait être établie d'après des sources nombreuses et variées de tous les niveaux de gouvernement et, en particulier, des documents historiques.

Dès 1953, Lewis Heck de l'ancien organisme appelé U.S. Coast and Geodetic Survey réclamait un répertoire national qui serait aussi complet que possible, dont le matériel serait puisé dans une grande variété de sources et qui renfermerait "suffisamment de détails pour permettre de repérer rapidement n'importe quel nom et qui donnerait les variantes des noms dans les cas litigieux".¹

A partir du milieu des années 50 jusque dans les années 70, de nombreuses tentatives furent faites, à l'intérieur comme à l'extérieur du gouvernement, pour établir des listes de noms exhaustives (certaines numériques); cependant, les produits obtenus manquaient de précision et étaient nettement incomplets, ce qui réduisait de beaucoup leur utilité. Dans bien des cas, on s'efforça de constituer des fichiers numériques sur lesquels on pourrait exécuter des opérations élémentaires de traitement de données, mais on ne possédait pas encore la technologie voulue pour pousser plus loin le développement et le traitement de fichiers de noms géographiques.

Divers utilisateurs avaient besoin d'un répertoire de noms géographiques unique, objectif et assez complet, qui devait contenir des renseignements fondamentaux ou essentiels sur chaque nom et chaque entité nommée. En raison de la nature et de l'ampleur de la tâche, la plupart des utilisateurs se tournèrent vers le gouvernement fédéral pour qu'on les aide à constituer et à tenir à jour un répertoire pilote, sur ordinateur, des 3 à 5 millions de toponymes des États-Unis.

L'USGS accepta en 1979 d'entreprendre un travail de ce genre dans le cadre du programme cartographique national et de participer au programme de normalisation nationale de l'U.S. Board on Geographic Names. L'USGS, qui avait expérimenté le traitement automatisé des noms géographiques depuis le début des années 70, constitua un prototype de base de données. L'amélioration des techniques d'extraction de données et de stockage de données auxiliaires rendit possible la mise en oeuvre de la NGNDB (base nationale de données sur les noms géographiques). Grâce à d'autres perfectionnements, on put développer et mettre en oeuvre un système complètement automatisé qui a pour nom GNIS, "Geographic Names Information System". Celui-ci comprend des bases de données, un logiciel d'extraction de données et des procédures; il se veut à la fois un répertoire officiel de noms géographiques et un outil de recherche toponymique.

Le système informatisé peut répondre à un large éventail de besoins, qu'il s'agisse d'informations ou de programmes:

- . aide à la cartographie
- . référence à des normes
- . base géographique de fichiers spéciaux
- . normalisation nationale
- . recherche toponymique et autre
- . applications numériques
- . inventaire et contrôle des noms géographiques
- . production de répertoires
- . publications spéciales

L'information contenue dans le système peut être restituée, organisée, analysée et manipulée pour divers usages à caractère général ou spécialisé. La base de données la plus importante et la plus volumineuse est la National Geographic Names Data Base, qui contient un fichier distinct pour chaque État, territoire et région isolée des États-Unis, ainsi que divers fichiers thématiques. Les autres bases de données du GNIS concernent:

- . les noms figurant sur les cartes topographiques nationales
- . les références et bibliographies
- . le Board of Geographic Names

RENSEIGNEMENTS DE BASE

Le GNIS et la NGNDB sont efficaces dans la mesure

1 Heck, L. (1953): "The problem of a national gazetteer", *Names*, vol. 1, n° 4, décembre, p. 233-238.

où les données qu'ils renferment sont complètes et exactes. En se fondant sur la recherche et l'expérience, on a retenu certains éléments ou catégories de données qui constituent un minimum obligatoire pour chaque nom ou chaque enregistrement de la base de données:

- l'orthographe du nom au long officiel, légal ou normalisé
- la désignation du type d'entité géographique
- le nom et l'indicatif de l'État et du comté
- le nom et l'indicatif de la carte topographique
- la position et l'étendue de l'entité nommée selon ses coordonnées géographiques
- le statut officiel du nom et son application
- les variantes ou formes secondaires du nom
- la référence bibliographique de la source d'où le nom a été tiré

Il y a aussi d'autres éléments d'information que l'on compile actuellement ou que l'on songe à introduire dans la base de données, par exemple l'altitude, la population, la description du lieu, l'origine et l'histoire du nom, la référence cadastrale, la référence à la propriété des régions administratives, ainsi que les indicatifs de la zone de recensement, de la zone postale et de la zone fluviale. D'autres catégories seront ajoutées au besoin.

COMPILATION DE LA BASE DE DONNÉES

Pour définir les éléments de données nécessaires et construire la base de données, il a fallu diviser le travail en plusieurs étapes, tâches et phases. La principale tâche consiste à recueillir les noms et l'information connexe à introduire dans la base de données. Les sources écrites d'information toponymique sont nombreuses et variées; y figurent notamment des cartes générales et cartes marines plus ou moins récentes, ainsi que des textes. Pour constituer la base de données, on ne retient que l'information publiée.

C'est sans aucun doute sur les cartes topographiques publiées par l'USGS que les noms géographiques actuels des États-Unis et de ses territoires sont les plus nombreux et les plus uniformément répartis. Une étude faite en 1970 par l'USGS montre que les cartes topographiques à grande échelle et à échelle intermédiaire contiennent entre 50 et 70 % de tous les noms actuellement en usage.

Il était donc logique de diviser le travail en deux phases distinctes:

- Phase I: le dépouillement toponymique des cartes topographiques de l'USGS;
- Phase II: le dépouillement toponymique des autres sources.

Phase I

L'USGS a entrepris le travail préliminaire de la phase I en 1976, à l'époque où les cartes topographiques à grande échelle couvraient plus de 70 % de la superficie du pays. A titre pilote, les noms figurant sur les cartes des États du Kansas et du Colorado ont été codés sous une forme assimilable par l'ordinateur. A l'automne de 1978, après une évaluation favorable du projet pilote, un contrat a été accordé pour faire de même avec les autres États et territoires; le travail

a été achevé en janvier 1982. On a vérifié tout le travail pour s'assurer qu'il était complet, exact et conforme aux procédures établies par la Branch of Geographic Names, qui appartenait à la Geological Survey's National Mapping Division. Le taux d'erreur était d'environ 2 %.

La phase I de la construction de la base de données a donc consisté en:

- le codage systématique de la plupart des noms et des renseignements connexes tirés des cartes topographiques, État par État;
- la correction et la mise à jour de ces données, en les comparant aux fichiers officiels du BGN.

Phase II

Pour rendre la base de données aussi complète que possible, il a fallu y ajouter les noms des entités qui ne figuraient pas sur les cartes topographiques mais que l'on pouvait tirer de diverses sources, notamment les documents historiques.

La phase II consiste donc essentiellement, pour chaque État, à dépouiller systématiquement des documents choisis pour créer des enregistrements toponymiques à ajouter à ceux qui avaient été introduits dans la base de données durant la phase précédente et à recueillir les variantes de tous les noms mémorisés. Ce travail doit être achevé avant la publication du volume du National Gazetteer of the United States of America de l'État ou du territoire considéré.

Les données recueillies dans le cadre de la phase I ne comprenaient pas toutes les entités nommées qui apparaissaient sur les cartes topographiques. Tantôt, on avait jugé que certains genres d'entités, qui étaient vastes ou dispersés, seraient plus faciles à manipuler pendant une phase ultérieure de la compilation; tantôt, l'information avait déjà été recueillie et figurait dans des listes spéciales ou sous forme numérique. Dans la phase II, il a donc fallu coder les noms qui avaient été omis délibérément dans la phase I, puis trouver et enregistrer les noms et l'information connexe qui n'apparaissaient pas sur les cartes topographiques.

Comme je l'ai mentionné précédemment, les cartes topographiques d'indiquent pas tous les noms géographiques actuellement en usage. Beaucoup de noms à usage local ou publiés étaient inconnus des cartographes au moment de l'établissement des cartes. Souvent, les noms désignant des entités de moindre importance n'étaient pas indiqués, faute de symboles cartographiques ou par souci de clarté. Aussi, en règle générale, les cartes n'indiquent pas les noms historiques ou désuets, les variantes et les autres noms qui ont pu être déjà utilisés pour désigner des lieux, des entités ou des régions. Des instructions complètes concernant la saisie, la documentation et le codage des noms géographiques dans le cadre de la phase II ont été publiées dans *The National Geographic Names Data Base: Phase II Instructions*.²

2 Orth, D.J. et Payne, R.L. (1984): *The national geographic names data base: phase II instructions*, U.S. Geological Survey Open File Report 84-036.

CRÉATION ET MISE EN OEUVRE DU SYSTÈME

Un des aspects les plus importants d'un système automatisé de noms géographiques est de permettre l'emmagasinage et l'extraction de l'information sans idée préconçue. Sur le plan fonctionnel, un système d'extraction de données sur les noms géographiques doit avoir cette qualité plus que toute autre. Ensuite, il doit avoir une grande rapidité d'exécution, assurer un bon cloisonnement des données, restituer les données sous diverses formes et être peu coûteux. Il doit aussi être facile à apprendre et à utiliser pour une grande variété d'utilisateurs. On doit pouvoir y ajouter facilement des données et des catégories de données et le relier à d'autres logiciels capables d'exécuter sur l'information des manipulations spécialisées et complexes.

Il aurait fallu beaucoup de temps, d'efforts et d'argent au personnel affecté au GNIS pour mettre au point un logiciel servant expressément à l'extraction de noms géographiques. L'USGS disposait déjà d'un programme qui offrait pratiquement tous les services jugés nécessaires à la mise en oeuvre d'un système toponymique automatisé. C'est pourquoi on a adopté le programme GIPSY³. Ce logiciel permet à l'utilisateur de communiquer directement avec la base de données en anglais, dans un langage clair qui évite toute ambiguïté. Une compétence et une formation en manipulation des données et une connaissance élémentaire des systèmes d'information suffisent pour soumettre les données à des recherches réflexives et introspectives. Les divers fichiers contenus dans un système automatisé de noms géographiques devraient être accessibles simultanément de plusieurs endroits, notamment depuis les bureaux d'information du gouvernement fédéral et des administrations des États et, éventuellement, au moyen d'ordinateurs individuels. Les programmes utilitaires GIPSY permettent d'ajouter des données

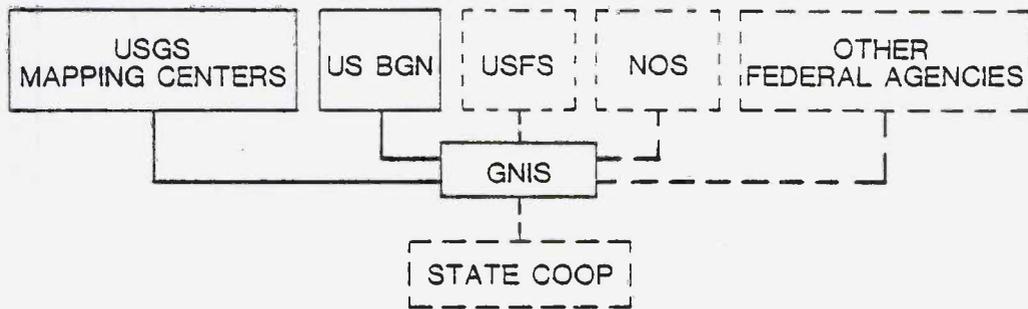
3 Le programme GIPSY[®] a été créé par le département des systèmes d'information de l'Université de l'Oklahoma à Norman, dans l'Oklahoma. Le GNIS est pour l'instant équipé de la version 1.0 de ce logiciel, mais on expérimente actuellement la version 2.0.

et des catégories de données presque indéfiniment. Il n'y a pas non plus de limite à la longueur des données, ce qui constitue une caractéristique indispensable d'un système toponymique automatisé.

ENVIRONNEMENT INFORMATIQUE

Le GNIS est utilisé dans le cadre d'une unité centrale de traitement, ce qui signifie qu'il fonctionne en multitraitement et utilise une mémoire virtuelle. Un répertoire permet d'identifier des entrées valides à l'intérieur d'un enregistrement; l'extraction des données comporte d'adressage dynamique d'un fichier temporaire. Il est possible de créer un index ou un fichier inverse dans le GNIS, mais la recherche répétitive n'est généralement pas employée dans un système toponymique automatisé. Le GNIS est accessible en tout temps, en dialogue, au moyen d'une liaison téléphonique. Il présente aux utilisateurs des menus faciles à utiliser, qui rendent possible une extraction de données en mode conversationnel (en temps réel) ou différé (par lots); à maints endroits, on peut aussi exécuter des copies sur bandes. Beaucoup d'utilisateurs sont prêts à se servir du système après avoir lu le GNIS Users Guide⁴ et en suivant les exemples étape par étape. Une courte formation peut également être dispensée à ceux qui s'intéressent à la structure des données et aux techniques d'extraction avancées. Ceux qui ne sont pas équipés pour consulter eux-mêmes le système peuvent, moyennant un tarif nominal, obtenir des services de l'USGS, notamment en matière d'extraction de données, de livres, de microfiches et de bandes magnétiques. L'accès au GNIS sera élargi grâce au réseau d'information géoscientifique (Earth Sciences Information Network) de l'USGS. Celui-ci consiste en une collection de programmes permettant aux utilisateurs de communiquer directement avec une variété de bases de données géoscientifiques qui résident sur différents ordinateurs centraux. L'usage de l'ESIN est pour l'instant limité

4 Payne, R.L. (1984): Geographic Names Information System Users Guide, U.S. Geological Survey Open File Report 84-551.



_____ INDIQUE UNE OPÉRATION OPÉRATIONNELLE

- - - - INDIQUE UNE PROCÉDURE DE DÉVELOPPEMENT

- USGS = US Geological Survey
- BGN = Board on Geographic Names
- USFS = U.S. Forest Service
- NOS = National Ocean Service

Figure 1 Le programme actuel de maintenance du Geographic Names Information System (GNIS)

aux bureaux d'information de l'USGS, mais on compte l'étendre à d'autres utilisateurs.

MAINTENANCE DU SYSTÈME

Un des aspects importants de l'exploitation d'un système d'information réside dans le programme et le calendrier de maintenance. Le programme de maintenance du GNIS n'est pas complètement prêt; cependant, la base du programme est établie et partiellement opérationnelle (voir Figure 1). C'est la National Mapping Division de l'USGS qui s'occupe de la maintenance du système. Chaque centre régional de cartographie compile et structure les nouvelles données et les corrections à des données existantes; celles-ci sont transmises électroniquement au personnel affecté au GNIS qui les soumet à une vérification finale avant de les introduire dans le système. L'U.S. Board of Geographic Names transmet des données directement au personnel du GNIS après avoir résolu des cas litigieux ainsi que des problèmes de normalisation nationale. Grâce à plusieurs vérifications et confrontations, on fait en sorte que les utilisateurs pourront extraire et utiliser des données complètes, exactes et bien protégées. La maintenance est un processus continu; le système est suffisamment souple pour nous permettre d'apporter aux données toute correction jugée nécessaire.

Si tous les utilisateurs peuvent extraire des données, seul le personnel du GNIS a la possibilité de modifier, d'ajouter ou de supprimer de l'information.

Le programme de maintenance fonctionne présentement dans les centres régionaux de la National Mapping Division de l'U.S. Board on Geographic Names; on songe à en étendre l'usage à d'autres organismes fédéraux et peut-être même à certains organismes d'État.

SOMMAIRE

Le GNIS est le fruit d'un long travail auquel ont pris part divers groupes en vue d'établir un répertoire unique, objectif et exhaustif des noms géographiques des États-Unis. Les résultats des recherches de plusieurs experts dans le domaine de la toponymie ont été mis à contribution pour développer et mettre en oeuvre un système toponymique automatisé. Celui-ci est conçu pour répondre aux besoins d'une large variété d'utilisateurs, à tous les niveaux de l'administration et dans le secteur privé, qui exécutent des recherches et s'appliquent à résoudre des problèmes d'application. En outre, le système facilite l'exécution des programmes nationaux de normalisation et de cartographie. Soumis à un programme de maintenance continu, il crée une variété de produits et de services.

TERMIUM III

John Carey*

Le Bureau des traductions met à la disposition de ses traducteurs et de ses clients une banque de terminologie informatisée, depuis plus de neuf ans déjà. Ce système, à ses débuts, portait le nom de TERMIUM I et fut élaboré à l'Université de Montréal en 1970. Vint ensuite sa prise en charge par le Bureau des traductions qui l'adapte à ses propres besoins (Terminum II).

Il devint bientôt clair que ce système de deuxième génération, aux possibilités techniques limitées, ne correspondait plus aux besoins d'un réseau d'utilisateurs en pleine expansion autant à l'étranger qu'au Canada. Par conséquent, de 1982 à 1985, le Secrétariat d'État a développé son propre système de troisième génération

(TERMIUM III) à partir du progiciel BASIS et de la technologie Télidon qui permet d'utiliser des menus et des graphiques de couleur. Il a aussi fait l'acquisition d'un ordinateur afin de mieux répondre aux besoins de la Banque et de faciliter l'accès à l'information.

Après la conversion des données à ce nouveau système, la Banque compte environ 750 000 fiches, ce qui représente quelque 1 800 000 termes.

Terminum III se compose d'une base de données linguistiques et d'une base de données documentaires. La base de données linguistiques à son tour se subdivise en quatre fonds:

- 1) fonds terminologique
- 2) fonds multilingue
- 3) fonds de difficultés de traduction
- 4) fonds d'appellations

C'est ce dernier fonds qui nous intéresse le plus aujourd'hui.

* John Carey, Agent, Service à la clientèle, Gestion du réseau, Direction de la terminologie, Bureau des traductions, Secrétariat d'État.

Il regroupe les fiches dont les vedettes sont des noms propres. Les types les plus courants d'appellations sont les appellations commerciales, les toponymes, les noms d'organismes, de comités et de programmes ainsi que les titres de documents et les noms officiels de lois et de règlements.

À l'heure actuelle, la banque compte environ 6 100 fiches dans le domaine des noms géographiques (ce qui comprend les toponymes). L'interrogation se fait à partir de menus qui sont en anglais ou en français selon les besoins de l'utilisateur. Voici un exemple d'une fiche: (voir Figure 1).

La fiche comprend plusieurs blocs d'information tels que: le bloc vedette, le bloc justification, le bloc sources et le bloc auteur. La fiche porte également la mention du fonds de données où elle a été classée ainsi que toute autre observation que l'auteur de la fiche juge utile d'ajouter.

Les terminaux donnant accès à TERMIUM servent

non seulement 1 200 traducteurs du Bureau, mais encore l'ensemble de la Fonction publique fédérale, et plusieurs gouvernements provinciaux, ainsi que des universités à Moncton, Montréal, Toronto, Ottawa, Kingston, Québec et Sudbury, divers instituts de recherche, des entreprises de traduction et des organismes étrangers tels que le Commissariat général de la langue française à Paris, le siège des Nations Unies à New York, le Fonds monétaire international à Washington et l'Organisation maritime internationale à Londres.

Gérée sur le principe d'une coopérative, la Banque devient la mémoire commune de tous les utilisateurs car ceux-ci ne sont pas simplement des clients achetant un service mais aussi des fournisseurs qui contribuent à l'alimentation du fonds commun en y apportant des nouvelles fiches ou en complétant les données existantes.

Tout comme la langue elle-même et la terminologie qu'elle emploie, la Banque est en constante évolution. Il s'agit d'une entreprise collective dont le succès repose sur la collaboration de ses nombreux partenaires.

```

                                ** Appellations - **
DOMAINE(S)                      : *Toponymes (Canada)
EN                               FR
Kicking Horse Pass*a           col Kicking Horse*b
OBS*Mountain pass, Canadian Rockies, on boundary bet. SE British Columbia and Banff National Park in Alberta.*a
OBS*col des Montagnes Rocheuses, à la frontière de l'Alberta et de la Colombie britannique; 1 628 m.*b
SOURCES
-----
a*WEGEO*1969***556
b*3DFX*1985
```

Figure 1

★★★★★★★★★★

RAPPORT DU SERVICE HYDROGRAPHIQUE DU CANADA SUR L'AUTOMATISATION
DES NOMS D'ENTITÉS SOUS-MARINES

M. Colin Bromfield* a fait un rapport succinct sur le système automatisé utilisé par le Service hydrographique du Canada (SHC) pour produire le Répertoire des noms d'entités sous-marines (1983). En 1982, les fiches relatives à quelque 3 500 noms d'entités sous-marines ont été versées dans un ordinateur PDP 11 par l'intermédiaire d'un programme écrit pour le progiciel DATATRIEVE, avant d'être transférées à un VAX 780. DATATRIEVE n'est pas un système de gestion de base de données mais un simple système d'interrogation élémentaire. Ce système, piloté par un menu, fonctionne en mode interactif en posant à l'opérateur inexpérimenté des questions pour

l'aider à accomplir certaines tâches telles que créer, corriger ou annuler des fiches contenues dans cette petite base de données. DATATRIEVE sert également à vérifier une grande partie des entrées de données. La représentation en 8 bits selon le format ASCII permet d'indiquer les accents sur les mots français. Pour mettre sur pied son système, le SHC s'est basé sur la BNDT, mais a modifié la présentation des fiches pour qu'elle réponde mieux à la consignation des données relatives aux noms d'entités sous-marines. La première édition du répertoire a paru en 1983. Actuellement, les travaux portent sur l'ajout sur les fiches de renseignements qui compléteront les données d'origine et d'historique des attributs. M. Bromfield a indiqué que la base de données du SHC peut être reliée à d'autres systèmes et que son contenu peut être restructuré de façon à être compatible avec un système de plus grande capacité ou plus perfectionné.

* Colin Bromfield, Analyste/Programmeur, Service hydrographique du Canada.

STATISTIQUE CANADA: ACTIVITÉS PORTANT SUR
LES NOMS GÉOGRAPHIQUES

Les représentants de Statistique Canada ont fait de brefs exposés afin de décrire les problèmes et les questions d'ordre toponymique rencontrés au cours de leurs activités.

Monsieur Henry A. Puderer*, a d'abord donné un aperçu général des services fournis par la Division de la géographie afin d'appuyer la collecte, la compilation et la diffusion de données statistiques par Statistique Canada. Ces services ou fonctions sont identifiés à la Figure 1; la Figure 2 montre les services géographiques fournis pour le Recensement du Canada de 1986.

M. Puderer a ensuite orienté son exposé sur

les questions toponymiques qui touchent les activités de ses services, en l'occurrence (a) la représentation de noms géographiques en haut de casse et en bas de casse, de noms comportant des accents, dans des dossiers, des cartes et des publications de Statistique Canada; et (b) les différences entre les méthodes de classification des noms de lieux utilisés par la Division de la géographie et ÉMR.

Dans le premier cas, M. Puderer a fortement recommandé de poursuivre l'élaboration de normes relatives à la représentation mécanique des accents et des procédures afin de faciliter au plus haut point l'échange d'informations toponymiques.

Dans le deuxième cas, M. Puderer a expliqué que les différences dans les méthodes de classification des noms de lieux utilisées par ÉMR et Statistique Canada nuisent actuellement aux efforts de Statistique Canada qui cherche à utiliser le même modèle de dossiers de

* Henry A. Puderer, chef de la Section de la délimitation spatiale et de l'analyse, Division de la géographie, Statistique Canada.

- Les six fonctions principales de la
Division de la Géographie**
1. Développement des concepts. Délimitation des unités géostatistiques.
 2. Cartographie pour la collecte du recensement.
 3. Fichier géographique principal de recensement.
 4. Géocodage.
 5. Produits cartographiques et tabulaires.
 6. Chiffres de population.

Figure 1

noms de lieux que celui de la BNDT d'EMR. Même s'il importe que Statistique Canada utilise le même modèle qu'EMR afin d'assurer que l'on trouve dans ses publications les noms officiels de lieux approuvés par le CPCNG, il est également essentiel qu'EMR ait facilement accès aux données de Statistique Canada au moment de préparer, par exemple, le Répertoire géographique du Canada et l'Atlas national du Canada. Statistique Canada continuera d'étudier cette question de similitude des dossiers en collaboration avec EMR.

Les observations de M. Joel Yan** ont surtout porté sur les questions de l'information de Statistique Canada relativement aux noms de rues et sur le placement automatique des noms sur les cartes.

** M. Joel Yan, chef du centre de la méthodologie de la Sous-Division de la géocartographie, Statistique Canada.

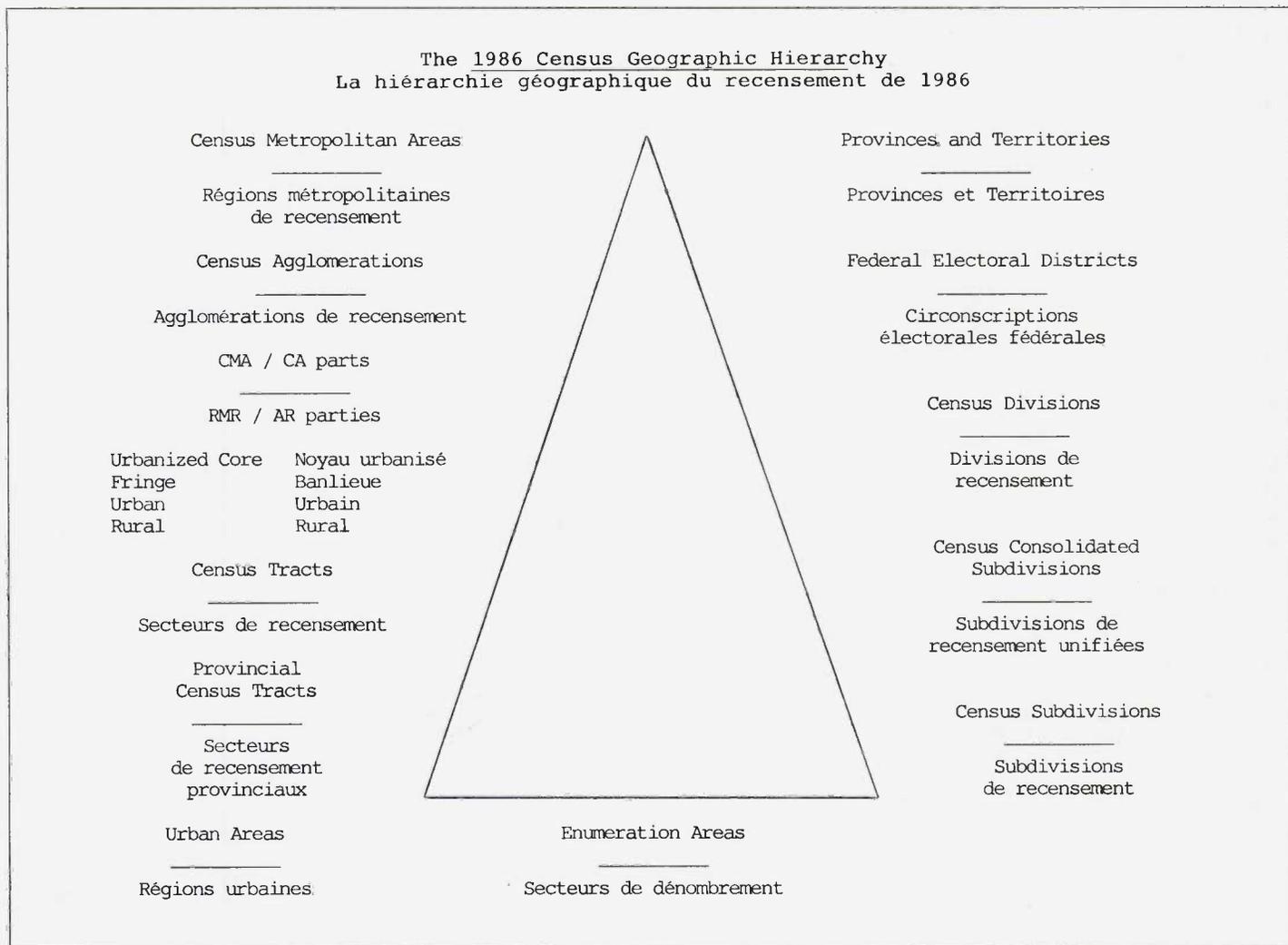


Figure 2

Statistiques Canada tient à jour des fichiers numériques sur les réseaux de rues afin d'offrir un système de sortie de données sur de petites zones de grands centres urbains (Fichier principal de la région (FPR). La couverture actuelle d'environ 200 municipalités représente près de 56 % de la population canadienne. Les fichiers comprennent, entre autres données, les noms de toutes les entités comme les rues, les routes, les voies ferrées, les réseaux hydrographiques et les limites principales. Les noms sont compilés à partir des renseignements contenus dans les documents des municipalités. Ces fichiers sont également utilisés pour la production de cartes automatisées à partir des données du recensement

(Figure 3) ainsi que de cartes automatisées à traits simples utilisées pour différentes applications par les secteurs privé et public (Figure 4).

Plusieurs questions ont été soulevées au sujet de la présentation des noms notamment:

- (a) la représentation de caractères spéciaux comme les accents;
- (b) le formatage (mise en forme) des noms; et
- (c) les usages et la vérification des noms de rues.

Le placement automatisé de textes sur des cartes



MODEL CT34PS

12/05/85 14 14 14 SCALE 3:00

Figure 3 Carte de base, secteur de recensement (SR)

Dès que le FPR est séparé par SR, les rues à double-lignes et les numéros d'ilot sont générés. Puis, le texte et le placement d'adresse est fait pour chaque fichier de SR. Le tracé obtenu est reproduit et utilisé par les représentants du recensement (énumérateurs) pour assurer une couverture ordonnée et complète de la distribution et de sa récupération des questionnaires du Recensement 1986. Environ 1,100 des 3,800 cartes de SR seront produites de façon automatisée pour 1986.

Le système a été développé par la sous-division de Géocartographie basé sur des entrées géographiques (Fichier Principal de Région et limites des secteurs de recensement) maintenues par la division de Géographie.



Figure 4 Carte pour la répartition des ambulances

(Partie du plan MH47 21/08/85 14 12 29 ÉCHELLE 20000)

Cette carte de référence à l'échelle 1:20,000 est parmi les premières d'un ensemble de 108 cartes produites de façon automatisée par Statistique Canada. Celle-ci fait partie d'un projet entrepris par le groupe de données Trillium afin de fournir au Ministère de la Santé de l'Ontario un système de cartographie et de répertoire de rues pour le service de répartition des ambulances du centre Halton-Peel. Les cartes, compatibles avec les cartes de base produites par le Ministère des Ressources naturelles de l'Ontario, feront partie d'un atlas contenant une série de cartes à être utilisées par les répartiteurs d'ambulances et leur personnel.

de rues urbaines est déterminé en calculant la longueur des lignes des textes et en trouvant sur la carte des segments de rues ou une paire de segments adjacents dont la longueur suffit pour permettre l'inscription du nom. On n'a pas considéré qu'il pouvait y avoir chevauchement sur d'autres lignes d'un texte; chaque

nom était placé sur l'élément linéaire auquel il correspondait. Il a fallu procéder au repositionnement interactif des textes ce qui a nécessité en moyenne 15 minutes par carte de secteur de recensement (Figure 3) et 6 heures par feuille de carte de 5 km sur 10 km (Figure 4).

Le placement de textes en polygone sur des cartes de municipalités aux fins de référence (voir les noms en majuscules à la Figure 5) repose sur les dimensions minimales du rectangle du polygone. Les lignes du texte étaient interrompues à un espace blanc approprié ou au moyen d'un trait d'union afin de donner à l'ensemble la forme approximative d'un rectangle et de placer im-

plicitement le texte au centre d'un polygone. Il n'y a pas eu vérification de cas possible de chevauchement d'autres lignes du texte. Cette méthode a donné de bons résultats lorsque utilisée sur des cartes de régions démographiques éparses (c'est-à-dire sur de grands polygones) et de moins bons résultats lorsqu'il était question de régions à plus forte densité démographique.

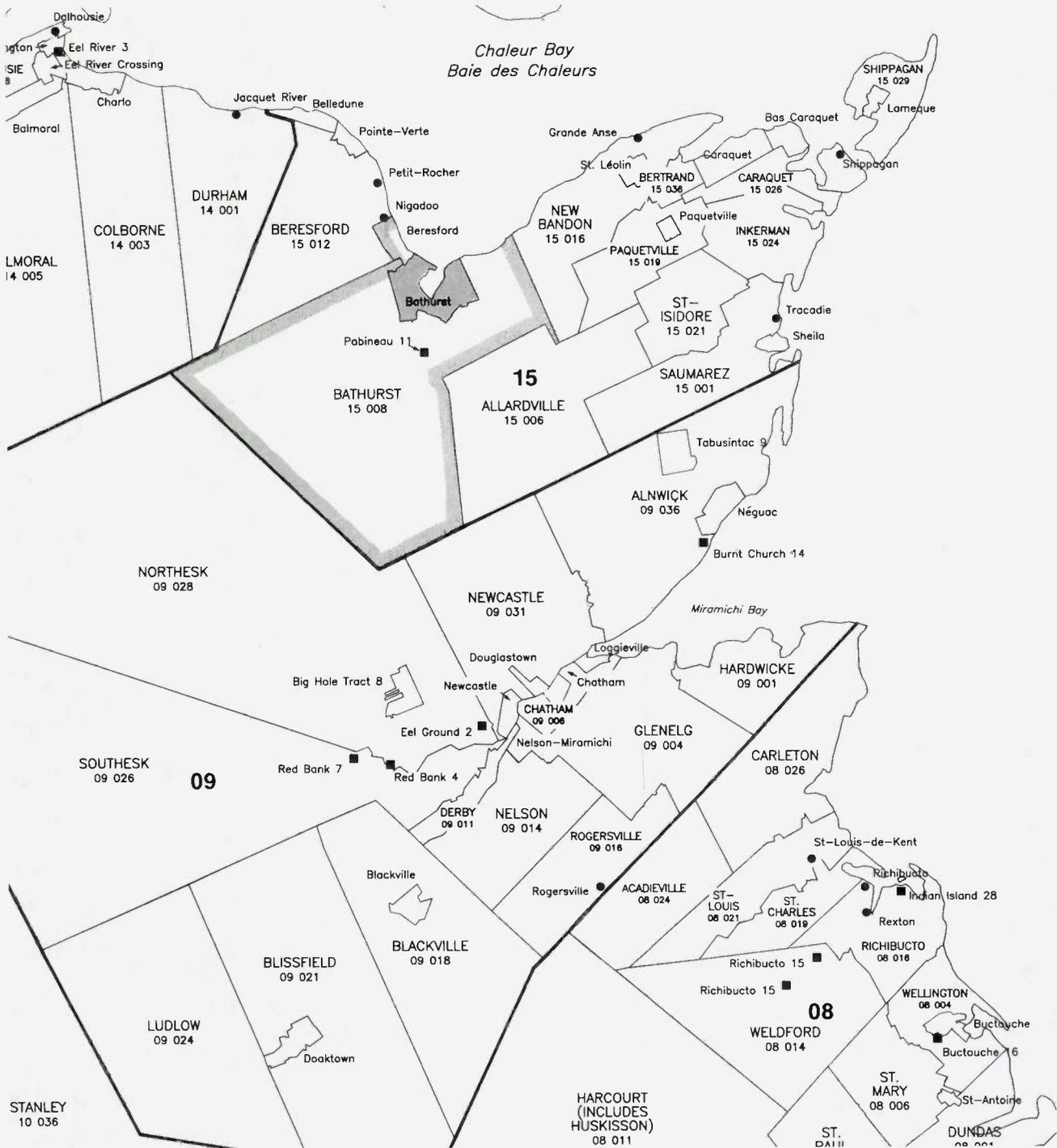


Figure 5 Une partie de la carte de référence du Nouveau-Brunswick montrant les divisions de recensement, les subdivisions unifiées et les subdivisions de 1981, avec la classification à sept chiffres (Classification géographique type - CGT)

ÉCHANGE EN PANEL ET COMMENTAIRE



Président:

Jean-Paul Drolet Président, Comité permanent canadien des noms géographiques, ÉMR

Richard J. Darley Senior Associate Chief, Cartography, National Geographic Society, Washington

Participants du forum:

John J.S. Thompson Directeur adjoint, Cartographie géographique, Division des services de géographie, ÉMR

Michael M. Munro Chef, Section de toponymie, Cartographie géographique, Division des services de géographie, ÉMR

Jeffrey R. Ball Spécialiste de la toponymie cartographique, Ministère des ressources naturelles de l'Ontario, Toronto

Gerald F. Holm Toponymist, Surveys and Mapping Branch, Department of Natural Resources, Winnipeg

Jean Poirier Adjoint au président, Commission de toponymie du Québec, Québec

Clément Nolette Étudiant gradué, département de géographie, Université Laval, Québec

Donald Orth Chief, Branch of Geographic Names, National Mapping Division, U.S. Geological Survey and Executive Secretary, Washington

Roger L. Payne Chief, Geographic Names Information Section, National Mapping Division, U.S. Geological Survey, Washington

Participants de l'auditoire:

Joel Yan Chef du centre de la méthodologie de la Sous-Division de la géo-cartographie, Statistique Canada, Ottawa

Norman Drummond Associate Professor, Department of Geography, McGill University, Montréal

Michael Smart Secrétaire exécutif, La Commission de toponymie de l'Ontario, Ministère des ressources naturelles de l'Ontario, Toronto

David Arthur Secretary, Saskatchewan Geographic Names Board, Regina

Douglas W. Kirk Ingénieur des systèmes hydro-métriques, Division des relevés hydrologiques du Canada, Direction des ressources en eau, Environnement Canada, Ottawa

William C. Wonders Professor, Department of Geography, University of Alberta, Edmonton

Marion O. Harris Consultant, Bell Communications Research, Madison, N.J.

Léo La Brie Traducteur, Bureau des traductions, Secrétariat d'État, Ottawa



Une variété de points et de questions est soulevée. Dans la mesure du possible, les points sont groupés et sont présentés sous les en-têtes suivants:

- (1) Le contenu et les limites de contenu des bases de données toponymiques
- (2) La coordination des bases de données de noms géographiques avec d'autres bases de données et avec la cartographie numérique
- (3) Les utilisateurs et leurs exigences en matière d'information sur les noms géographiques
- (4) Les problèmes particuliers de l'automatisation des noms géographiques:
 - (a) odonymes
 - (b) caractères accentués

- (5) La prononciation des noms géographiques et les répertoires concis
- (6) Recommandations

- (1) LE CONTENU ET LES LIMITES DE CONTENU DES BASES DE DONNÉES TOPONYMIQUES

M. Thompson explique que la Direction des levés et de la cartographie avait pris une décision réfléchie dès le début de l'automatisation de la Base nationale de données toponymiques (BNDT) de créer un système toponymique, plutôt qu'un système d'information géographique.

Des attributs tels que la population, l'étendue, les conditions climatiques, le revenu annuel moyen, les altitudes, etc., qui ne sont pas strictement de caractère toponymique n'ont pas été inclus. M. Thompson indique cependant que plus tard les données toponymiques pourraient se fusionner avec d'autres travaux en cours afin de créer un système d'information géographique de plus grande envergure.

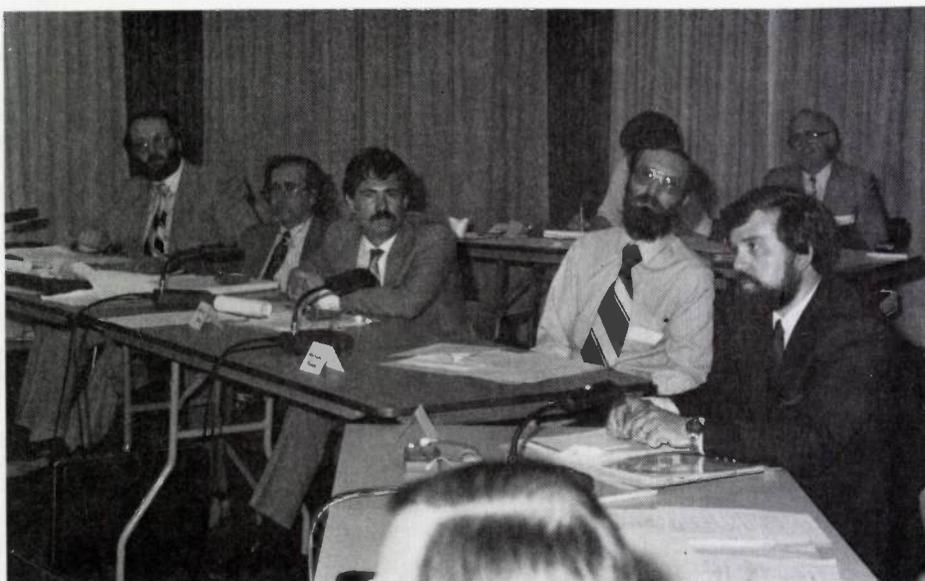
M. Orth signale qu'aux États-Unis chaque nom géographique est étroitement lié à d'autres données reliées à un élément géographique précis. Ainsi, la méthode américaine n'a pas été limitative, mais a été orientée vers l'expansion afin d'inclure des informations que d'autres domaines n'avaient pas encore pris en charge, la cartographie automatisée par exemple. M. Orth mentionne qu'il y a bien sûr une limite au genre de données qu'on accepte d'inclure. En ce qui concerne par exemple le débit des cours d'eau, on procéderait en sens inverse: le système d'information sur les noms géographiques (Geographic Names Information System: GNIS) fournirait l'information de base (ou un sous-ensemble) que la division des ressources en eau de l'USGS utiliserait et fusionnerait dans leurs fichiers. M. Payne souligne que le GNIS demeure un système automatisé de noms plutôt qu'un système d'information géographique. On est maintenant en train d'ajouter certains éléments ou catégories de données (par exemple la population), en réponse aux demandes des utilisateurs. Dans le cas des chiffres de population, l'US Bureau of Census (qui est représenté au sein de l'USBGN) accepte la responsabilité de l'intégrité des données ajoutées au système.

M. Darley explique que la National Geographic Society avait élaboré un programme interne de vérification de l'orthographe, mais avait trouvé qu'il n'était pas rentable sauf si d'autres fonctions pouvaient être exécutées avec le fichier des noms. Une fois que les chiffres de

population et d'autres attributs sont ajoutés, il sera possible d'extraire automatiquement l'ensemble des écritures cartographiques du fichier. Ces économies en heures-personnes aideront à compenser le coût de construction du système.

M. Holm fait le point sur les progrès réalisés avec l'informatisation des fichiers des noms au Manitoba pendant les 10 dernières années. Le VAX 11/750, acheté en 1983 est utilisé avec un dérouleur de bande DEC TE 16 et un logiciel DECUS (DEC User-Society) qui comporte un éditeur, un vérificateur d'orthographe et un programme de manipulation de bande. Dix-neuf terminaux et deux imprimantes sont maintenant exploités au sein de la direction; deux terminaux et une imprimante peuvent traiter et imprimer des caractères internationaux. Une bande de fichiers du Manitoba a été introduite dans la BNDT à Ottawa et a été comparée avec la série de données utilisée pour produire le répertoire annuel du Manitoba. On a produit de cette façon une liste de correction d'erreurs qui servira à corriger les fichiers de la BNDT. M. Holm signale qu'en collaboration avec le personnel des Levés et de la Cartographie au 615, rue Booth, il vient d'accéder pour la première fois de l'extérieur au système informatisé des noms géographiques du Manitoba. Le système du Manitoba est maintenant compatible avec ceux de la plupart des directions provinciales des Levés et de la Cartographie au Canada qui sont munis de VAX ou DEC PDP 11. Actuellement, le Manitoba utilise le système pour préparer un supplément cumulatif du volume manitobain du "Répertoire géographique du Canada"; il s'agit d'un procédé rapide qui produit dans quelques heures une version sur papier.

M. Poirier signale qu'en raison d'un manque de fonds, la Commission de toponymie ne s'est pas servie du système ONOMA. La Commission espère maintenant disposer de micro-



PARTICIPANTS DU FORUM:

(De gauche à droite): Clément Nolette, Jean Poirier, Jeffrey R. Ball, Michael R. Munro, John J.S. Thompson

ordinateurs et de moyens d'infographie pour l'information toponymique dans deux ou trois domaines de service.

M. Arthur dit que le gouvernement de la Saskatchewan s'intéresse à une base automatisée de données toponymiques. Il espère que dans quelques années il y aura des fonds suffisants pour l'acquisition du matériel et la liaison à la BNDT. Dans l'intervalle, le vieux système sur fiches devra suffire.

MM. Orth et Payne font ressortir qu'aux États-Unis aucun État n'avait de bases de données automatisées pour les noms géographiques avant l'informatisation du GNIS. Ensuite, cependant, plusieurs États ont mis sur pied des bases automatisées de données toponymiques en incorporant les données du GNIS dans leurs propres systèmes. En ce qui concerne les noms géographiques et leurs applications, tous les organismes fédéraux sont tenus de se conformer aux décisions de l'USBGN.

(2) LA COORDINATION DES BASES DE DONNÉES DE NOMS GÉOGRAPHIQUES AVEC D'AUTRES BASES DE DONNÉES ET AVEC LA CARTOGRAPHIE NUMÉRIQUE

M. Orth donne des précisions sur certaines particularités du GNIS qui ont été mises au point afin de l'intégrer au programme de cartographie numérique. Par exemple, les coordonnées géographiques ont été enregistrées à la seconde la plus proche (avec une tolérance de ± 5 secondes) et dans le cas des cours d'eau les coordonnées de la source et de l'embouchure ont été enregistrées. M. Payne ajoute que l'échelle ne pose pas de problème. La série des fonds de cartes utilisée pour le programme de cartographie numérique est à 1/24 000 ou 1/25 000. La numérisation de l'information toponymique s'est faite à partir de copies papier à la plus grande échelle disponible à ce moment-là; ainsi 75 % de l'information de la phase I ont été numérisés à partir de cartes topographiques à 1/24 000 et les 25 % restants à partir de coupures à 1/62 500. M. Payne explique qu'un programme de recherches vient d'être amorcé pour établir une liaison entre le GNIS et la base de données cartographiques numériques; les problèmes (y compris celui de la numérisation à partir d'une copie papier plutôt que des dessins originaux) seront déterminés pendant la phase initiale des travaux de recherche.

M. Ball indique que l'Ontario est en train de commencer la numérisation du système de référence géographique à l'aide d'un prototype. Il faut que l'information sur les positions que l'on introduit dans les fichiers des noms géographiques et partant dans la base future de données automatisée soit compatible; certains problèmes se posent par exemple en ce qui concerne les différences entre l'utilisation d'un système de référence UTM et l'utilisation de coordonnées géographiques.

M. Yan dit qu'aux États-Unis, on s'applique à établir un comité fédéral de cartographie numérique chargé de coordonner toutes les activités fédérales relatives au système d'information géographique et cartographique. M. Thompson explique que des représentants des divers ministères fédéraux commencent à aborder les problèmes de l'incompatibilité dans l'échange de données numériques entre les organismes fédéraux canadiens. Le comité, qui a tenu sa première réunion à Ottawa, compte s'occuper des problèmes suscités par la multitude de données numériques et prendre des mesures telles que des échanges de personnel et de normes proposées portant précisément sur l'information géographique et la production cartographique. On espère que le comité se penchera aussi

sur le traitement des toponymes, des accents et des définitions normalisées de génériques.

En ce qui concerne la liaison avec la cartographie numérique, M. Orth pense qu'éventuellement au lieu de consulter des cartes sur support papier, les particuliers pourront afficher l'information qu'ils désirent sur un écran cathodique dans leur bureau, chez eux ou dans leur voiture. Afin de satisfaire aux besoins des utilisateurs qui choisissent des champs de données pour une région précise, la base de données des noms géographiques doit être en mesure de fournir l'information de référence nécessaire pour ce genre de cartographie future. M. Orth mentionne qu'en effet, à l'USGS, les fichiers de noms géographiques sont plus avancés que beaucoup d'autres aspects de la cartographie numérique.

M. Drummond s'enquiert du placement automatique des noms et de la façon dont on pourrait coordonner une banque de noms géographiques avec une banque de données sur les ressources afin d'obtenir une intégration exacte par exemple des noms et des unités géologiques et de leurs codes.

M. Thompson dit qu'en dehors de certains exemples tels que le choix de caractères utilisés dans les documents de cartographie du recensement (Statistique Canada), il existe une difficulté réelle en ce qui concerne la bonne dimension et le placement correctement orienté des noms même en ce qui concerne un but unique, sans parler de l'intégration avec un deuxième groupe d'information relative aux noms. Il se peut qu'il soit nécessaire de changer les normes cartographiques actuelles; cela ne figure cependant pas parmi les priorités actuelles.

M. Orth est d'avis qu'on s'aventure maintenant dans des problèmes cartographiques qui ne sont pas nécessairement liés à la normalisation ou à l'automatisation des noms. Le placement automatique des noms se trouve encore à ses débuts, mais ce ne sera pas long avant que tous les éléments d'une carte soient identifiés et placés automatiquement. Il ajoute que, théoriquement, les géologues américains ne sont pas autorisés à utiliser des noms géographiques non officiels dans les rapports; les noms géologiques, bien qu'ils soient fondés sur les noms géographiques, n'appartiennent pas aux données du GNIS.

(3) LES UTILISATEURS ET LEURS EXIGENCES EN MATIÈRE D'INFORMATION SUR LES NOMS GÉOGRAPHIQUES

M. Payne fait le point sur les utilisateurs des données du GNIS. Il dit que des dossiers relativement complets sont tenus sur les utilisateurs du système; en général les utilisateurs proviennent dans une proportion de 60 % de l'industrie privée, 35 % du gouvernement et 5 % d'autres secteurs. Le système représente l'ensemble de la société allant des sociétés pétrolières et des entreprises de camionnage aux particuliers faisant des recherches généalogiques. M. Payne dit qu'il est normal d'avoir environ 10 à 20 commandes par semaine. Le genre d'information demandé varie; cependant, l'une des demandes les plus communes concerne toutes les montagnes situées dans un État donné et dont l'altitude dépasse un chiffre donné. Les autres types de demandes comprennent les noms de toutes les agglomérations à une altitude supérieure à 6000 pieds (demandés par une entreprise chargée de la distribution d'un projet donné); des renseignements pour planifier les itinéraires de livraison dans un rayon de 100 milles de Houston; les cimetières se trouvant dans un rayon donné; tous les lacs et réservoirs dans une région donnée. Il suffit

d'une fraction du temps et du coût qu'il aurait fallu pour obtenir les mêmes données en vérifiant les cartes pour répondre à ces demandes. M. Payne relève que l'on ne connaît probablement pas la plus grande tranche d'utilisation des données du GNIS. Des commandes d'achat de bandes numériques sont constamment remplies par l'entremise d'un bureau central. L'information est transférée dans la base de données des acheteurs qui l'appliquent ensuite à leurs besoins.

M. Wonders demande à M. Ball des précisions sur la base d'information qui a servi aux conclusions de son exposé. Les 17 réponses semblent évaluer les réactions des utilisateurs gouvernementaux au sondage. Il désire connaître s'il existe un fondement statistique pour l'évaluation de l'utilisation des données sur les noms géographiques de l'Ontario par les organismes non gouvernementaux, y compris sociétés, établissements d'enseignement, écrivains, particuliers, etc. Il trouve aussi que les résultats du sondage sont nécessairement déformés lorsqu'aucun facteur de pondération n'a été utilisé pour tenir compte du nombre de demandes provenant d'une même source.

M. Ball répond que les 21 personnes interrogées avaient été retenues comme utilisateurs primaires et secondaires des données en se fondant sur d'autres statistiques; elles constituent le groupe-cible principal de l'étude sur l'utilisation des différents aspects des fichiers écrits et graphiques.

M. Darley demande si l'Ontario avait l'intention de définir des éléments (zonaux) sous forme de polygones et les rivières comme des éléments linéaires marqués d'un grand nombre de coordonnées le long de leurs cours. M. Ball répond que seul un nombre minimal de coordonnées serait enregistré: traditionnellement ce serait le centroïde pour un polygone, l'embouchure et la source pour les cours d'eau, plus les coordonnées du début et de la fin sur une coupure particulière. Il trouve que l'idée d'entrer un plus grand nombre de coordonnées afin de pouvoir définir la superficie et mettre au point le placement automatique des écritures, dépasse nettement la capacité d'un système toponymique.

M. Drummond dit qu'il voudrait savoir quelles sont les observations des ministères chargés des ressources, comment ces utilisateurs choisissent ou éliminent les noms dont ils ont besoin pour leurs cartes thématiques et s'ils ont leur propre fichier de toponymes, qui permettrait par exemple d'associer la présence de certains types de roches avec certains noms géographiques.

M. Ball dit qu'en Ontario, le programme de cartographie de base de l'Ontario (Ontario Basic Mapping: OBM) fixe ses propres critères pour l'utilisation des toponymes et que les autres utilisateurs dans la province sont libres de le faire aussi. Il indique qu'habituellement les géologues demandent que des noms soient attribués à des éléments sans nom afin de rendre leurs rapports plus clairs. M. Smart dit que les fichiers de la Commission des noms géographiques de l'Ontario fournissent des renseignements sur les noms officiels, par opposition aux noms non officiels, utilisés éventuellement par des équipes d'arpentage. M. Ball explique que la production de cartes de nomenclature géographique constitue un projet de l'OBM. Chaque carte porte la nomenclature approuvée jusqu'à une date donnée. Les nouvelles décisions seront incorporées ensuite manuellement dans le "cronaflex", sur la carte elle-même et sur la liste des noms dans la marge de la carte. Un exemplaire du fond de carte mis à jour serait ensuite expédié à tous ceux inscrits

sur la liste des destinataires.

M. Holm fait le point sur les cartes toponymiques supplémentaires distribuées en novembre 1984, selon la formule de récupération des coûts, aux utilisateurs dans les ministères manitobains. Ces cartes à 1/250 000 portant en surcharge la mise à jour des noms sont destinées particulièrement au personnel oeuvrant dans le domaine des ressources dans la partie septentrionale de la province. Un exemplaire a été mis à la disposition du grand public dans le bureau provincial des cartes à Winnipeg. M. Holm ajoute que la plupart des demandes ne concernent pas l'encombrement des noms sur les cartes, mais plutôt les nombreux éléments sans nom. On peut satisfaire aux demandes de données provenant des sociétés des ressources en leur fournissant une image-mémoire de l'ensemble du fichier du répertoire annuel provincial, ce qui est peu coûteux et peut se faire à court préavis. Les toponymes non officiels utilisés dans le passé par les explorateurs et les géologues par exemple, peuvent se trouver dans les fichiers à titre de variantes; il n'existe cependant pas une liste complète de ces noms.

M. Thompson explique que la Direction des levés et de la cartographie est chargée de l'Atlas national du Canada ainsi que de l'exploitation de la BNDT. Une façon de répondre aux problèmes d'encombrement cartographique et toponymique consiste à choisir un certain nombre de fonds de cartes sur lesquels les géographes peuvent travailler. Des fonds de cartes formatés d'avance pourraient comprendre des réseaux de drainage en traits fins, moyens ou grossiers, des symboles d'agglomérations, etc.; l'utilisateur pourrait ensuite choisir entre différents menus pour obtenir les combinaisons souhaitées d'information. Cette façon de procéder enlève à l'utilisateur la liberté absolue, mais impose une certaine rigueur et aboutit à une certaine cohérence du produit, tout en réduisant les erreurs dans la compilation de données telles que les noms de lieux. On résiste aux pressions d'inclure un plus grand nombre de toponymes pour servir des demandes isolées de cartographie thématique, mais la sélection des toponymes pour les fonds de carte aurait pour objectif d'inclure les principales agglomérations, stations climatologiques, etc.

M. Kirk demande plusieurs questions concernant l'obtention de données en mode interactif et la disponibilité de systèmes faciles à utiliser au Canada. Il est d'avis que les bases de données aboutissent souvent à des versions qui se ressemblent énormément; il demande si l'on a examiné la possibilité d'utiliser le même programme pour extraire des données des différentes bases provinciales. Aux États-Unis, où le GNIS est facile à utiliser, M. Kirk se demande si le nombre des demandes de renseignements avait diminué et si les différents utilisateurs ont accès à différents fichiers.

M. Thompson répond que la BNDT est facile à utiliser pour la personne qui connaît le système et qui est à l'aise avec celui-ci; la documentation est relativement longue et malheureusement, étant donné les particularités actuelles du matériel et du logiciel, il n'est pas inhabituel de perdre un fichier complet ou le travail d'un après-midi à un terminal interactif. Bien que le système ne s'approche nullement de l'envergure du système américain, il est probablement plus complexe et ne se prête pas facilement à un accès par ordinateur personnel. Puisque les provinces élaborent leurs propres systèmes, chacune a des priorités différentes en ce qui concerne les ressources financières et humaines. Au Canada, le gouvernement fédéral ne pouvait pas attendre que chaque province ait mis en place une base de données sur les

noms géographiques avant d'élaborer et de coordonner un système central. Dans son rôle cartographique, la Direction des levés et de la cartographie devait utiliser les toponymes fournis par le secrétariat du CPCNG et devait amorcer les travaux.

M. Payne souligne que l'expression "facile à utiliser" a un sens différent pour différentes personnes. En principe, à l'USGS, le bureau des noms géographiques est orienté sur la recherche plutôt que sur les services aux utilisateurs. A partir d'un mode d'exploitation qui n'était pas au début "facile à utiliser", le système est devenu depuis 1980 d'accès de plus en plus facile (à l'aide de menus), les services aux utilisateurs étant confiés à une vingtaine de points d'information où le personnel affecté aux demandes de renseignements a reçu une courte formation pour apprendre les commandes du système. La partie difficile est de comprendre la nature des données dans le système, afin de bien répondre aux questions des clients; c'est le domaine de formation sur lequel on se concentre maintenant. M. Orth ajoute qu'en essayant de rendre le système plus facile à utiliser, on a éliminé les codes dans la mesure du possible; la recherche des codes introduit un délai et éventuellement des possibilités d'erreur. M. Payne dit que les codes sont retenus pour le traitement numérique et sont par conséquent utilisés sur les bandes magnétiques; mais dans le cas des réponses aux demandes des utilisateurs, on interroge le système en utilisant des caractères en clair sans code.

M. La Brie demande quelles mesures il faudrait prendre si un traducteur voulait connaître par exemple tous les noms comportant un certain générique; comment faudrait-il procéder et quel en serait le coût?

M. Thompson demande d'adresser les questions précises au secrétariat du CPCNG. La capacité du système permet d'extraire tous les codes génériques par province et ensuite plus précisément en fonction de paramètres, comme une coupure de carte du SNRC ou dans une région géographique définie selon des latitudes et des longitudes. L'on s'efforce de travailler selon une formule de récupération des coûts. Dans le cas d'autres ministères et organismes gouvernementaux, si le personnel consacre peu de temps à la demande, on fait payer seulement les coûts directs de la main-d'oeuvre et des systèmes et non les frais généraux ou les frais attribuables à la supervision. Par exemple, le coût d'une bande contenant tous les toponymes du Manitoba qui a été envoyée au gouvernement provincial, était de plusieurs centaines de dollars. En principe, la première demande d'un dénombrement de lacs, par exemple, coûterait quelques centaines de dollars; tandis que la deuxième demande concernant les mêmes renseignements coûterait beaucoup moins cher parce que le fichier existerait déjà.

M. Payne explique que les coûts de données fournies par le GNIS dépendent de l'étendue de la région faisant l'objet de la recherche. Une recherche concernant tous les étangs dans un État coûterait entre 1,50 \$ et 10 \$, tandis qu'une recherche des fichiers pour l'ensemble des États-Unis (plus de 2 millions d'inscriptions) coûterait entre 50 \$ et 60 \$ US. Graduellement, au cours des cinq dernières années les demandes des clients ont été transmises au National Cartographic Information Center, qui a des moyens d'accès en mode interactif au GNIS. Le système a été conçu pour ce genre de demande de renseignements et s'en tire très bien. L'on tient des registres des demandes reçues. M. Orth ajoute que l'on peut fournir une liste de tous les genres de génériques apparaissant dans un polygone définissable. L'on peut

aussi porter sur des cartes ces renseignements, ce qui a été fait en ce qui concerne la distribution des différents génériques aux États-Unis.

M. Poirier explique que tel que recommandé par les Nations Unies, la Commission de toponymie fait la distinction entre le terme générique, par exemple "crique" dans Crique Honorius, et "l'entité", qui concerne une classification plus large de la terminologie générique, qui est dans ce cas "ruisseau".

(4) LES PROBLÈMES PARTICULIERS DE L'AUTOMATISATION DES NOMS GÉOGRAPHIQUES

(a) Odonymes

M. Poirier fait le point sur les odonymes (noms de rues) au Québec. Le "Guide toponymique" a été produit en 1979 afin de fournir des règles pour l'écriture des noms géographiques, y compris les noms de rues. Actuellement la Commission de toponymie partage avec les municipalités la responsabilité de rendre officiels les noms des rues, et la Commission estime qu'il y en a un total de quelque 100 000 qui doivent être approuvés officiellement. Récemment la Commission a commencé à entrer les noms de rues dans la base de données et certains ont déjà été soumis à la Commission pour approbation. L'on espère que dans un an il sera possible d'extraire de l'ordinateur les odonymes selon la municipalité, et dans deux ans d'établir la liste de tous les toponymes dans une municipalité donnée. A l'heure actuelle la Commission traverse une période de transition, et fait des expériences avec l'utilisation de micro-ordinateurs. Récemment, aux fins d'une plus grande précision, les coordonnées des deux extrémités d'une rue ont été enregistrées. La procédure n'est pas encore appliquée aux cours d'eau puisqu'il y a plus de problèmes à déterminer exactement la source d'une rivière lorsqu'elle se trouve dans un marécage ou un bassin.

Bien qu'il soit possible d'identifier les autoroutes par numéro, seuls leurs noms sont enregistrés dans les dossiers officiels et la Commission essaiera d'indiquer les coordonnées géographiques dans les différentes municipalités que les autoroutes traversent.

M. Orth indique que l'entrée d'odonymes dans le GNIS serait une activité de la phase III. Pour deux raisons, ces données n'ont pas été introduites auparavant: a) jusqu'à maintenant, il a été très difficile de recueillir ces renseignements de façon cohérente dans l'ensemble du pays; les efforts déployés par l'US Bureau of Census et l'US Postal Service sont en train de changer la situation maintenant; b) l'on n'a pas résolu comment identifier, dans le système, l'étendue des éléments. L'on espère maintenant lier cette étape au programme de cartographie numérique, une fois que l'on commencera à numériser les rues et les routes; à ce moment, il sera possible d'identifier des points précis le long d'un élément et de donner un nom à la série de points.

(b) Caractères accentués

M. Darley signale que jusqu'à maintenant la National Geographic Society examine l'extraction automatique des écritures mais non le placement

automatique, pour un système qui aura tous les accents mondiaux, qui les affichera sur un écran et pourra les imprimer sur une petite imprimante. Une fois que l'orthographe et les accents sont justes, les bandes produites pour la typographie donneront des résultats plus justes. En exploitant le grand ordinateur central IBM, et un bloc de programmes d'infographie acheté d'IBM mais légèrement modifié pour un terminal muni d'un clavier normal, la société a créé des alphabets comportant toutes les lettres accentuées. Il est possible d'afficher maintenant l'information sur l'écran et de donner des instructions au matériel typographique sur les caractères qui doivent être munis d'accents, qu'il s'agisse de majuscules ou de minuscules. M. Darley dit que ce bloc de programmes standard d'infographie de l'IBM a été adapté aux besoins de la National Geographic Society et fonctionne parfaitement.

Les premiers toponymes introduits dans le système de la National Geographic Society seront les noms de l'Atlas mondial de la société; l'information sera mise à jour en même temps en utilisant l'orthographe officielle qui devrait figurer sur les cartes. L'un des problèmes est celui de déterminer comment leur magazine voudra utiliser les données dans le cas des publications destinées au public de langue anglaise. A cette fin, il faudra établir le catalogue des toponymes en précisant si la version se réfère à une carte, à un magazine ou à un livre. Dans le magazine, on n'aura pas besoin d'accents, mais dans les cartes les accents seraient utilisés selon le système de translittération. Les deux séries de toponymes doivent être introduites dans le système et étiquetées pour que le chercheur puisse extraire la version dont il a besoin. M. Darley souligne que les caractères sont créés à partir d'une matrice sur l'écran. Lors de la réception du bloc de programmes d'infographie, la société a dû élaborer et introduire chaque lettre et chaque configuration des accents possibles dans le système afin de créer le fichier.

M. Orth explique que l'USGS s'occupera bientôt de cette question d'accents. A l'heure actuelle, le service a les moyens nécessaires pour munir les toponymes français et espagnols des signes diacritiques, mais il est en train d'enregistrer les toponymes amérindiens, notamment les toponymes Navajo et Apache qui utilisent leur propre alphabet romain modifié, utilisant des obliques, des combinaisons de lettres et des apostrophes sur et sous les lettres. Ce sera un domaine très important et M. Orth espère que certaines sociétés commerciales réussiront à fournir des claviers adaptés à une grande variété de symboles peu familiers.

M. Yan exprime le souhait de donner suite à la question de l'étude et de l'utilisation de caractères spéciaux.

(5) LA PRONONCIATION DES NOMS GÉOGRAPHIQUES ET LES RÉPERTOIRES CONCIS

M. La Brie voudrait savoir si la prononciation serait fournie pour les 50 000 noms qui seront inclus dans le "Concise Gazetteer of the United States of America".

M. Orth répond que le Canada est en avance sur les États-Unis en ce qui concerne les normes de prononciation. Une prononciation normalisée est très difficile à établir;

par exemple, il y a au moins 15 prononciations locales de Chicago. Il signale que le répertoire concis a été élaboré en réponse à une recommandation du groupe d'experts des Nations Unies des noms géographiques au moment où un programme de répertoires nationaux pour le monde entier était planifié et que le nombre total d'inscriptions avait été attribué aux pays selon leur dimension et leur population. Les bibliothèques aux États-Unis préfèrent aussi un seul volume des lieux habités et des entités géographiques à titre d'ouvrage de référence pour qu'il ne soit pas nécessaire de consulter le répertoire de chaque État. Le "National Atlas of the United States of America" qui contient entre 40 000 et 50 000 noms géographiques, principalement de lieux habités auxquels on a ajouté les noms des principaux éléments physiques pour un total de 50 000, est une base de toponymes. On ne projette pas à l'heure actuelle de préparer un guide de prononciation.

M. La Brie fait remarquer que l'Autriche est le seul pays qui a produit un répertoire qui comprend des indications sur la prononciation.

M. Holm dit que le Manitoba avait reçu un certain nombre de demandes provenant des médias d'information concernant un guide de prononciation, notamment en ce qui concerne des toponymes amérindiens.

Mme Harris mentionne l'annuaire phonique automatisé utilisé par la Bell. La mauvaise prononciation des noms n'est pas réellement tenue pour un problème, dans la mesure où la prononciation des noms suit des règles et ressemble à ce que l'on s'attend et que les gens peuvent les comprendre, même si la prononciation n'est pas "correcte".

(6) RECOMMANDATIONS

Il est recommandé:

- (a) qu'au fur et à mesure que des normes pour l'automatisation des noms géographiques sont élaborées au Canada, elles soient publiées et diffusées, peut-être par l'entremise des Nations Unies, pour que d'autres puissent examiner la documentation et que toutes les autorités nationales puissent entamer la normalisation mondiale dans ce domaine;
- (b) que des efforts soient déployés pour coordonner les programmes et les besoins de Statistique Canada et de la Direction des levés et de la cartographie, ÉMR, en ce qui concerne l'information toponymique;
- (c) que des efforts soient déployés pour normaliser le traitement des lettres accentuées dans les banques de données toponymiques;
- (d) que les autorités en matière de toponymes considèrent la préparation de fonds de cartes toponymiques automatisés traduisant les décisions prises dans leur domaine de compétence et qu'elles élaborent des moyens pour informer les utilisateurs de l'existence des cartes, des modifications et des nouveaux toponymes;
- (e) que des lignes directrices proposées et des règles générales relatives aux conventions touchant aux odonymes soient recueillies et fassent l'objet d'échanges entre les autorités toponymiques et Statistique Canada.

FIRST SYMPOSIUM ON THE AUTOMATION OF GEOGRAPHICAL NAMES
PREMIER COLLOQUE SUR L'AUTOMATISATION DES NOMS GÉOGRAPHIQUES
OTTAWA, JUNE 4, 1980/LE 4 JUIN 1980
CANOMA, VOL. 6, NO. 1, JULY/JUILLET 1980

Contents/Sommaire:

Excerpts from the opening remarks of Jean-Paul Drolet/Extraits d'une allocution d'ouverture par Jean-Paul Drolet

Papers/Documents

File structures for handling toponymic data	James G. Linders
Automation of the National Toponymic Data Base	John J.S. Thompson
Système automatisé de gestion de données toponymiques	Pierre Barabé
Traitement des noms géographiques à l'aide de systèmes d'information automatisés à référence spatiale	C.B. Berthold et B.J. Rivard
Identification of rivers for stream-flow data	D.W. Kirk and W.J. Ozga
A national geographic-names data base	Donald J. Orth
Automation of foreign place names at the United States Defense Mapping Agency	William E. Opalski
Automation of geographical names in Michigan	Stewart A. Kingsbury

Panel discussion and commentary/Échange en panel et commentaire

Copies are available from CPCGN Secretariat, (EMR), 650-615 Booth Street, Ottawa,
K1A 0E9

Les textes sont disponibles au Secrétariat du CPCNG, (ÉMR), 650-615 rue Booth,
Ottawa, K1A 0E9