Énergie, Mines et Ressources Canada
Energy, Mines and Resources Canada

CANOMA

Vol. 6 – No. 1 July / juillet 1980

AUTOMATION OF GEOGRAPHICAL NAMES

AUTOMATISATION DES NOMS GÉOGRAPHIQUES

Nouvelles et commentaires concernant la Toponymie du Canada recueillis par le Secrétariat du Comité permanent canadien des noms géographiques

Publié par la Direction des levés et de la Cartographie, Energie, Mines et Ressources, Canada

News and Views concerning Canadian Toponymy compiled by the Secretariat of the Canadian Permanent Committee on Geographical Names

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

ISSN 0319-5228

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

Secretariat
Canadian Permanent Committee on Geographical Names
Room 124A, 615 Booth Street
Ottawa, Ontario, K1A 0E9

If you would like a copy of any article in the other official language please write to the Secretariat.

Toute information concernant CANOMA ou noms géographiques en général devrait être envoyée au:

Secrétariat
Comité permanent canadien des noms géographiques
Pièce 124A, 615, rue Booth
Ottawa, Ontario, K1A 0E9

Si vous désirez recevoir le texte d'un article dans l'autre langue officielle, veuillez écrire au Secrétariat.

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: Blair MacDougall, 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: Blair MacDougall, Centre d'Informatique, EMR)

CANOMA

Vol. 6 - No. 1 July 1980 Vol. 6 - No. 1 juillet 1980

SYMPOSIUM ON THE AUTOMATION OF GEOGRAPHICAL NAMES

COLLOQUE SUR L'AUTOMATISATION DES NOMS GÉOGRAPHIQUES

EXCERPTS FROM THE
OPENING REMARKS
OF
JEAN-PAUL DROLET*

I am certainly pleased that so many experts on geographical names have accepted the invitation of the Canadian Permanent Committee on Geographical Names to attend this automation symposium, and I believe that today's large attendance is also an indication of the strong interest in this subject. In addition to several representatives from various departments of the federal government, there are large contingents from both Ontario and Quebec, and representatives from New Brunswick, Newfoundland, Saskatchewan and Alberta. From the United States Geographic Board we have Mr. Orth, and from the United States Defense Mapping Agency we have Mr. Opalski. Specialists from the academic world include Dr. Linders from the University of Guelph, Dr. Kingsbury from Northern Michigan University and Mme Berthold and M. Rivard from Université de Sherbrooke.

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

EXTRAITS D'UNE
ALLOCUTION D'OUVERTURE
PAR
JEAN-PAUL DROLET*

Il m'est certainement agréable de constater que tant d'experts en toponymie ont bien voulu accepter l'invitation du Comité permanent canadien des noms géographiques à participer à ce colloque sur l'automatisation; par votre présence, vous manifestez également le très grand intérêt porté au sujet. Outre plusieurs représentants de divers ministères fédéraux, un grand nombre de délégués est venu du Québec et de l'Ontario de même que des représentants du Nouveau-Brunswick, de Terre-Neuve, de la Saskatchewan et de l'Alberta. Nous comptons également parmi nous M. Orth, du United States Geographic Board, et M. Opalski, du United States Defense Mapping Agency. D'autres spécialistes, du milieu universitaire, sont également présents. Citons notamment le professeur Linders de l'Université de Guelph, le professeur Kingsbury de la Northern Michigan University ainsi que Mme Berthold et M. Rivard de l'Université de Sherbrooke.

* Jean-Paul Drolet, président du Comité permanent canadien des noms géographiques.

We are honoured today by the presence of Mr. Rolf Böhme, Scientific Director of the Institute of Applied Geodesy in Frankfurt am Main, West Germany. Mr. Böhme is also the convenor of the United Nations Working Group on the Automation of Geographical Names.

I am very impressed by the range of topics we have before us this morning. The papers by Dr. Linders and Mr. Opalski related to the theoretical and developmental aspects of automated systems. The studies by Mr. Thompson, Mr. Orth and Mme Berthold and M. Rivard concern the implementation of systems. Dr. Kingsbury's paper is on the creation of a data bank of origin information, and Mr. Kirk's paper examines the problems faced by the user of automated systems who needs reliable and unambiguous geographical name information.

One person who is conspicuous by his absence today is my good friend, Henri Dorion. As the head of the Advisory Committee on Toponymy Research he had urged the convening of a symposium on geographical names for the past five years. He was eager to make the system being developed in Quebec compatible with the system being established in the federal Surveys and Mapping Branch so that mapmakers and other users seeking names information could fulfill their requirements without cumbersome procedures. Henri has just been appointed Quebec's Delegate General to Mexico, and, I understand, is presenting his credentials to the Mexican government today.

In convening this symposium I believe the Canadian Permanent Committee on Geographical Names is performing some of the important tasks for which it was established over 80 years ago: the responsibility for standardizing the use of geographical names throughout Canada, the function of bringing together personnel from the various provincial and federal names authorities to resolve problems of common interest, and the concern for the advancement of toponymy.

Some of the questions to which we have to address ourselves today are:

(1) How can we utilize modern computer technology in a way that will facilitate the recording of official geographical names and, concurrently, make such information available to a variety of users in useful formats?

(2) How can the various systems presently being developed be standardized so that information may be exchanged from one system to another?

(3) How can computer technology be used in order to improve the quality and availability of geographical names information at a time when spending restraints discourage any increase in manpower?

(4) How can the information on names be protected from change or alteration once its use has been authorized by the appropriate responsible authority?

(5) How can the use of names information in an automated system be assessed in order to repay the high costs of systems development and management, and to what extent should the users have to bear the cost?

M. Rolf Böhme, directeur scientifique de l'Institut de géodésie appliquée de Frankfurt am Main, Allemagne de l'Ouest, nous fait également honneur de sa présence. M. Böhme préside le groupe de travail des Nations Unies sur l'automatisation des répertoires de noms géographiques.

La diversité des thèmes dont nous devons traiter ce matin est très impressionnante. Les documents présentés par MM Linders et Opalski traitent des aspects théoriques et évolutifs des systèmes automatisés. Les études présentées par Mme Berthold et MM Thompson, Orth et Rivard portent sur la mise en place des systèmes. Le document du professeur Kingsbury concerne la création d'une banque de données sur l'origine des noms, et enfin le document de M. Kirk permet d'étudier les problèmes auxquels fait face l'utilisateur des systèmes automatisés en quête de sources de renseignement fiables et non ambiguës en matière de noms géographiques.

Une absence est aujourd'hui fort remarquée. Il s'agit de mon bon ami Henri Dorion qui, à titre de chef du comité consultatif de la recherche toponymique, insistait depuis cinq ans pour qu'un colloque du genre soit organisé. Il souhaitait adapter le système mis au point au Québec à celui qui a été mis au point à la Direction fédérale des levés et de la cartographie pour permettre à ceux qui sont chargés d'établir les cartes et aux autres utilisateurs à la recherche de données sur les noms géographiques d'accomplir leurs tâches sans formalités fastidieuses. Henri vient d'être nommé délégué général du Québec à Mexico; il doit justement présenter aujourd'hui ses lettres de créance au Gouvernement du Mexique.

En organisant ce colloque, le Comité permanent canadien des noms géographiques accomplit l'une des importantes missions qui lui a été confiée il y a plus de 80 ans, soit la responsabilité de normaliser l'utilisation des noms géographiques au Canada, de réunir le personnel spécialisé en noms géographiques des provinces et du gouvernement fédéral pour résoudre des problèmes communs et de faire évoluer la toponymie.

Certaines questions auxquelles nous devons essayer de répondre aujourd'hui sont les suivantes:

(1) Comment pouvons-nous utiliser la technique moderne de l'informatique de façon à faciliter l'enregistrement des noms géographiques officiels et par conséquent, mettre ces données au service de différents utilisateurs et ce, de façon pratique?

(2) Comment peut-on normaliser les différents systèmes actuellement mis au point pour pouvoir transférer les données d'un système à un autre?

(3) Comment utiliser la technique de l'informatique pour améliorer la qualité et favoriser une meilleure diffusion des données relatives aux noms géographiques à une époque où un régime d'austérité ne permet pas l'accroissement de la main-d'oeuvre?

(4) Comment peut-on éviter que les données sur les noms géographiques ne soient changées ou modifiées une fois qu'elles ont été acceptées par une source autorisée?

(5) Comment peut-on évaluer l'emploi de données sur les noms géographiques intégrées dans un système automatisé pour amortir les coûts élevés de la mise au point et de la gestion des systèmes, et dans quelle mesure les utilisateurs devraient-ils en assumer le coût?

I trust that some answers to these questions, and others that I have not mentioned, will be forthcoming today so that when we draw up conclusions and recommendations later this afternoon we will be able to take positive steps in coordinating the development and use of automated names systems.

J'espère que nous trouverons des réponses à ces questions ainsi qu'à d'autres que je n'ai pas mentionnées; nous pourrons ainsi tirer des conclusions et formuler des recommandations vers la fin de l'après-midi et adopter des solutions positives pour coordonner la mise en oeuvre et l'utilisation des systèmes automatisés.

******

FILE STRUCTURES FOR HANDLING TOPONYMIC DATA

James G. Linders*

ABSTRACT. The development of toponymic data bases is discussed in terms of existing digital technology. A review of the essentials of information processing and file structuring notions is followed by a specific information structure for handling large ordered files. The system proposed has been implemented at the University of Guelph, and the results have been found to be consistent with the performance of real time systems. The extensions of this system are also discussed in terms of other functions associated with the operation of toponymic data bases.

RÉSUMÉ. La mise sur pied de banques de données toponymiques est discutée en fonction des techniques existantes en matière de données numériques. Un rappel des fondements du traitement des données et des notions d'établissement de dossiers est suivi par la description d'un système d'information conçu spécialement pour la gestion des dossiers ordonnés et volumineux. Le système proposé a été mis en pratique à l'université de Guelph, où l'on a jugé les résultats compatibles avec le fonctionnement des systèmes en temps réel. Il est également fait mention des possibilités d'adaptation de ce système à d'autres fonctions associées à l'exploitation de banques de données toponymiques.

INTRODUCTION

The importance of toponymic data bases can hardly be overemphasized in a world where information is of increasing importance in terms of timeliness, precision and accuracy. The development of large information networks, reaching down even to the level of an individual home within the wired city context, provides even greater motivation for migrating current information bases to digital technology. Whereas much of this information has in the past been available in gazetteers and indexes, it is clear there are other more appealing avenues within the current information resolution.

The advantages to moving to a digital base are obvious. First, the information is more readily available on demand in a current and more reliable format. Secondly, the ease of updating and the relative efficiency of the process result in justifiable cost benefits that are hard to refute. Thirdly, the distribution of the information in alternative formats to those provided by gazetteers can provide many side benefits. For example, the information may be formatted and selected according to pre-defined criteria so that only those aspects which are meaningful to the user need be provided on demand. This latter alternative is particularly attractive within an information network environment. In this context the quality of the product can be controlled to the user's requirements: everything from a simple printout on a computer terminal to a high quality photo typeset script for the preparation of reports or other documents as required for circulation and publication. In Marshall MacLuhan's famous quotation, "the medium is the message", can be found a rationalization for an information utility providing for geographical information in a manner meaningful to the end user. Toponymic data bases provide yet only one facet of an information system environment with other similar file systems providing complementary information related either to a theme or a statistical aggregation according to the user needs.

* James G. Linders, Chairman, Department of Computing and Information Science, University of Guelph.

SYSTEM REQUIREMENTS

To appreciate the considerations implied in the development of a digital toponymic data base there is a need to have an understanding of the fundamentals of digital information processing. At the highest conceptual level the concern is primarily with structuring a system environment which can realize the specific user needs at a lower level. Considerable thought must go into the design and operation of the data base implementation. Perhaps one of the most important aspects of any systems development is the human interface describing both the ease of use and the clarity of function. It is appropriate, therefore, to review some of the relevant notions of information science which are applied in the development of toponymic data bases.

System Environment

The system environment is concerned with providing an operational and conceptual framework for realizing the practical implementation of a specific functional system. Traditionally this involves the identification of the type of computing system as well as its associated devices and their characteristics, which in turn dictate the expected performance of the implementation. Other aspects of the system environment which may not be so readily obvious are the operating systems under which the user application normally runs. The user's evaluation of any application for a system is usually in terms of quantifiable parameters such as the response time for obtaining results to queries, as well as the cost of executing specific identifiable tasks, such as making queries. Another identifiable parameter is the storage required to maintain a particular data file or data base as well as its associated costs. Unfortunately the cost of running an application is highly dependent upon the specific type of system environment under which the application has been implemented.

Applications Development

There are many options normally available to the implementor of new applications. Wherever possible an existing package should be used as opposed to developing new customized software unless the application has attributes or characteristics which necessitate the extra expenditure and time. Especially where the application is simple it is normally possible to find a generic piece of software which can be customized to a wide spectrum of related tasks. This does not negate the need for careful analysis of the requirement before embarking upon such developments because there are often hidden pitfalls from which recovery is both difficult and expensive. A package can only be used effectively when the analysis and implementation are consistent with the user needs.

The management of toponymic data represents an application which can be as elementary as a simple information retrieval capability based on structured files to a complex data management problem requiring access to a data base management system. For the former the requirement is for the orderly storage, management and retrieval of primitive information, whereas, for the latter the concern is primarily with providing information based on the encoding of primitive data types as well as the relationship they bear to each other. Essentially the former can be viewed as straight forward fact retrieval providing the information that is stored within the file system, while for the latter new information is inferred from the collection of available data within the associated file systems.

Information Needs

A basic requirement in developing a toponymic data base is to identify the various types of data and their organization within the associated file systems. As with most data processing applications the intended uses often dictate the corresponding system structure.

In any data system it is normal to differentiate between the logical structure of the information versus the physical structure or implementation. At the logical level information is organized in terms of collections of attributes to form larger groups. In terms of an implementation the attributes constitute fields which are grouped together according to some pre-defined format to constitute a record. The information flow in and out of the file system is at the record level. The records are organized within the file according to some pre-defined scheme. Normally one field is identified as a key field and the records are ordered on the collating sequence of this key field. If the file is structured on a particular key field this normally dictates the form of access to the records. In some cases it is possible to have multiple keys providing essentially varied access to the data records themselves.

The fields represent the encoding of data attributes or characteristics. Information may be stored either in character form or as arithmetic entities. Even though variable length fields are possible especially for character fields it is normally simpler to allow a maximum length for the particular field and pad to the right with blanks.

The primary concern in the development of any file system is the organization of the records within the file. Access to the records has been defined to be in terms of either their sequential position in the file or via key fields. The key is essentially a label which can be translated into an address to access a specific record directly. The organization of records in a serial manner requires that any particular record can only be accessed in a sequential manner. That is, a given record can be accessed only if its predecessor has been already passed. In a direct access structure a table gives the relationship between the key of the record and its physical address on direct access storage medium, such as disk. If the file is very large the table providing the relationship between the key field and the address can be potentially very large. In order to accommodate such large files it is normal to select certain key fields and store them into a table. This mechanism of building an index combines the advantages of direct access and sequential access. Essentially the address of specific keys, normally those beginning at a cylinder and those beginning for each track are defined in terms of a table. Having found the particular cylinder and subsequently the particular track on which the record resides a sequential search is then performed on the track in order to find the record being sought. This form of access is called index sequential and is commonly found within the file management software within any reasonably sophisticated operating system.

B-TREE STRUCTURES FOR A DATA BANK

The advantages of a direct access file can often be achieved within sequential files by using appropriate index structures. The index sequential method found within IBM's Data Management services represents an initial structuring of data in terms of index structures. A more sophisticated and improved method includes the building of index structures by means of B-Trees. The primary consideration is of course to reduce the average time required to access a record.

A B-Tree of order N is characterized by the following properties:

(1) Each node in the tree has no more than N immediate successors.

(2) Each node in the tree except for the root and the terminal nodes has at least N/2 immediate successors.

(3) The root has at least two immediate successors, but could have none.

(4) Each nonterminal node with K immediate successors has K-1 keys associated with it. Each key value in the nonterminal node delimits the extent of the keys in the immediately successive nodes of the tree.

(5) Terminal nodes can either contain the data records or pointers to the data records.

The basic rationale for developing B-Trees is to minimize the search times for any item. A variety of search techniques has evolved to permit a search of sequential files. These are inherently expensive, even for a binary search the average number of comparisons which must be made is of the order of $\log_2 N$. This would mean that for a disk file containing 100 000 records, at least seventeen disk accesses would be required for each search. Even in a sophisticated computing system, this would require several seconds before reasonable access could be assured for the user. The B-Tree search on the other hand is much more effective, in that a B-Tree tends to be short (that is, has very few levels) and wide (that is, there are very many nodes at each level). This is as opposed to a binary tree which tends to be very long (has many levels) and narrow (has few nodes at each level). Also, in a binary search only half of the remaining keys are eliminated from consideration at each comparison. Whereas in a B-Tree search, approximately (N-1)/N of the remaining keys are eliminated at each step. The maximum search length for B-Tree depends upon the

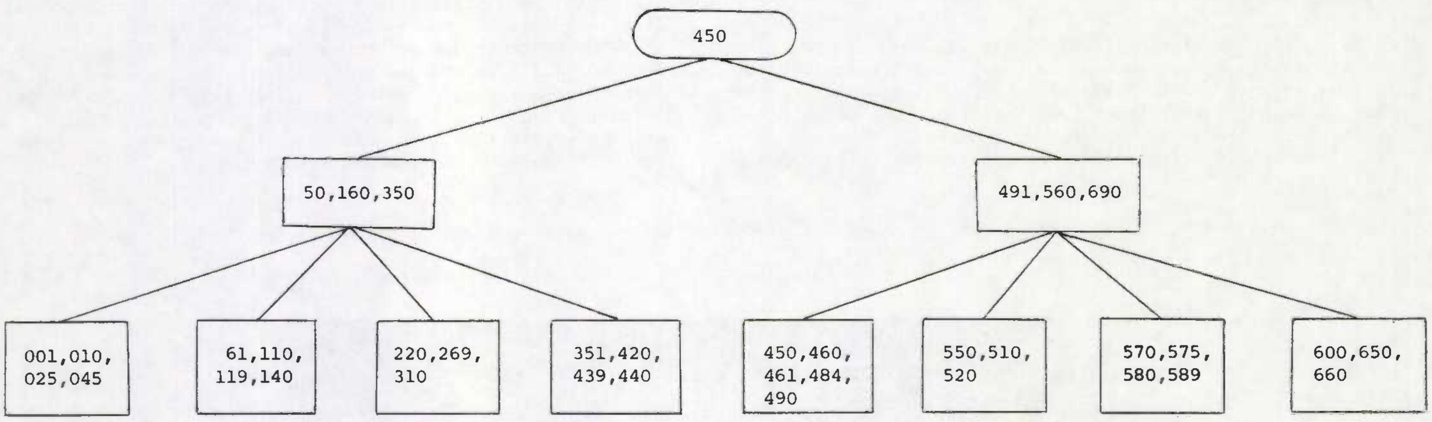

Figure 1 A B-Tree structure of order 5

Figure 1 shows a B-Tree structure of order 5. The root of the tree is represented by a key value of 450. All keys less than 450 would be found in the left hand side of the tree. All keys greater than 450 or equal to 450 would be found on the right hand side. The first hand sibling of the B-Tree contains a node with three values. The first value of 50 delimits the upper limit of keys found in the immediate successor to this node. The process would continue to any number of levels. It should be noted that all keys are unique and that the keys are found within the tree in ascending or descending sequence. When a new key is added a node would expand to accommodate the new key value unless the maximum size of the node was exceeded, at which point the node must be split into two new nodes. This in turn could cause a splitting at the next higher level, etc.

number of levels in the tree. For a B-Tree in which the terminal nodes are at level K+1 the maximum search length is K. If each B-Tree node contains N keys the number of terminal nodes is N+1 and the relationship between the search length and number of keys is defined by the inequality:

$$K \leq 1+\log_{\lceil \frac{n}{2} \rceil} \frac{N+1}{2}$$

This means that for a B-Tree of order 256 containing 100 000 records, the maximum search length is no greater than three.

In a practical implementation the size of N would be dictated by the buffer size and block size of records on the disk. Ideally each node should contain as many keys as possible.

IMPLEMENTATION

A B-Tree package has been developed, and written in a high-level programming language called C and running under the UNIX operating system at the University of Guelph. This package has been used to store a number of files including the *Populated Place Names* file as well as the *Physical Features* file of the Canada Gazetteer Atlas. Other files such as the *Control Survey* file for Ontario have also been structured under this B-Tree arrangement.

For the *Physical Features* file, the key used is feature name. Even though the feature name is not unique, a single entry is stored for each key. When the data record is located it is flagged as to whether it has multiple values or not. All records which have the same key value are chained and can be displayed on demand.

Normally only a single key is associated with each file, although this need not be the case. For the *Control Survey* file for Ontario, the file is structured with station number being the primary key. Since each station number is unique this presents no problems. Other indexes are also built to provide alternative access to the file, namely station name as well as position.

At this time no formal query language exists; however, this will be provided in the very near future to allow the user to interrogate the system through a generalized query language. At present, retrieval can only be obtained by keying in a known key and having the record returned online. The specific implementation referenced in this paper is on a PDP 11/34 mini computer and response is clearly within real time, namely less than a few seconds for any record in the relatively large files. Each file is approximately in the order of 20 000 records each.

New fields can be added in the files as required. When other key fields are necessary, then an appropriate B-Tree index must be created on each key.

Another program currently exists to allow the definition of a mask for defining the format and entering data into the file system. As records are entered, the B-Tree structure is searched to find the appropriate position in the B-Tree and the key is inserted. At the same time the record is stored in the next sequential position in the file and the position noted in the index structure of the B-Tree.

SUMMARY

This particular system was developed as a package for organizing and managing large data files for use within an online environment. It is intended to use this system within the georeferencing system environment, whereby many different user files will be available to a wide variety of users. As each file is brought to the system it is only necessary to create the appropriate index structure and maintain the index structure for all data changes in the file. The programs will remain in C, however, an attempt will be made to migrate them to another programming language such as FORTRAN to provide greater portability. As well some cosmetic changes to the system will simplify the user interface.

Photo: Mike Kelly, Information EMR

SYMPOSIUM SPEAKERS/ORATEURS DU COLLOQUE:

Seated at front tables (left to right)/Assis aux tables (de gauche à droite:
James Linders, Donald Orth, John J.S. Thompson, Pierre Barabé, Douglas Kirk.

AUTOMATION OF THE NATIONAL TOPONYMIC DATA BASE

John J.S. Thompson*

ABSTRACT. The Toponymy Unit, National Geographical Mapping Division, Energy, Mines and Resources, is responsible for the integrity and maintenance of the National Toponymic Data Base. An ever-increasing volume of geographical names data and a requirement to more rapidly process and disseminate these data necessitated the introduction of computer technology. A commercially available system, DATABOSS/2, was purchased and modified for application in the data base. Large volume encoding of Canada's toponymic records is now in progress, with completion scheduled for December 1981. System automation has permitted production of the first computer generated and type-set volume of the Gazetteer of Canada series. Resources that would be needed for the manual maintenance of ever increasing volumes of toponymic data will now be freed by automation for reassignment to the qualitative improvement of the National Toponymic Data Base.

RÉSUMÉ: La Section de toponymie, Division des cartes géographiques nationales, ministère de l'Énergie, des Mines et des Ressources, est responsable de l'intégrité et de la tenue de la base nationale de données toponymiques. Le volume toujours croissant de données sur les noms géographiques et le besoin de traiter et de diffuser ces dernières avec rapidité ont nécessité l'introduction de l'informatique. Un système commercial, le DATABOSS/2, a été acheté et modifié pour répondre à nos besoins. L'enregistrement d'un grand nombre de dossiers toponymiques sur le Canada est en cours et devrait être terminé en décembre 1981. Ce système a permis de produire et de composer entièrement par ordinateur le premier volume de la série des Répertoires géographiques du Canada. Les ressources qui sont normalement requises pour l'entretien manuel du volume sans cesse croissant de données toponymiques pourront maintenir servir, grâce à l'informatique, à l'amélioration qualitative de la base nationale de données toponymiques.

INTRODUCTION

The Toponymy Unit, National Geographical Mapping Division, of the Surveys and Mapping Branch, Department of Energy, Mines and Resources, is responsible for ensuring the availability of toponymic information, including names approved by the Canadian Permanent Committee on Geographical Names, to the federal government and to the general public. It is a task that departmental staff have been performing for many years. During those years, however, significant changes have taken place in the magnitude and nature of the task.

Some time ago it became obvious that, due to the ever-increasing volume of toponymic data, new techniques were required to ensure a responsible level of geographical name acquisition, storage and retrieval, both for federal mapping and in support of the endeavours of the CPCGN. The task of maintaining the data base was labour intensive and time consuming. Large staff commitments to clerical functions meant that resources were not available to undertake field research programs, even though such programs formed a fundamental part of the maintenance of the National Toponymic Data Base.

In 1977, the Geographical Services Directorate, Surveys and Mapping Branch, undertook an evaluation of both the operational effectiveness of its toponymic service and the quality of the toponymic data base. The evaluation indicated that the data base contained approximately 350 000 approved name records in primary files and an additional 650 000 names in secondary files. It was determined that information entered the system at a rate of approximately 2000 new approved names and 300 name revisions per month, and an overall rate of increase of approximately 20% per year. Although the Toponymy Unit was responding well to information demands for individual names, limited resources prevented it from responding efficiently to requests for large lists of names for federal map makers. These same limited resources also meant that the Unit was having difficulty in rapidly updating the data base upon receipts of batches of new information from the provinces.

The evaluation indicated that the increase in the volume of names information could no longer be handled without the introduction of appropriate new technology. It was noted that the consistent structure of the bulk of the names information lent itself to the automation of the storage, updating and retrieval functions which would otherwise require large amounts of clerical manpower. Furthermore, as various aspects of map production were making ever increasing use of digital technology, it was important to ensure that the toponymic files were compatible with these other related applications. Computer technology, it appeared, would allow the conversion of clerical manpower to the professional staff positions needed to carry out and supervise the fieldwork required for the qualitative improvement of the data base. At the same time, automation would facilitate the storage and retrieval of the data.

---

* John J.S. Thompson, Acting Head, Cartography Section, National Geographical Mapping Division, Surveys and Mapping Branch, EMR.

Following the 1977 evaluation, a contract was awarded to a national computer consulting firm for recommendations regarding the specific application of computer technology to the work of the Toponymy Unit. The consultants' final report, submitted in March 1978, indicated that the automation of the data base would:

(1) significantly increase capability to store and disseminate accurate toponymic information;

(2) reduce clerical operations;

(3) provide enhanced security for the data base;

(4) allow the automated creation of names lists and type for cartographic compilers;

(5) facilitate timely production of the various volumes of the Gazetteer of Canada series; and

(6) allow greater flexibility in the use of names data for geographical and historical research.

The consultants initially suggested that a totally custom designed data base management software package, running on a mini-computer system, would best fill our automation needs. The Directorate felt, however, that in view of the many proven and well documented commercial data base management systems available, it would be worthwhile for the consultants to carry out a review of these various commercial packages. Eventually, two such packages were recommended, one named DATABOSS/2 and the other named DATATRIEVE. Both packages were reviewed in depth, with evaluation results indicating that DATABOSS/2 apparently held the greater promise for our proposed application. Of considerable importance in selecting a package, was the adaptability of DATABOSS/2 for use on existing Branch computer equipment. On-site testing of the package was done at a computer facility in Halifax, Nova Scotia, with satisfactory results. Thus, in the fall of 1978, a decision was made to purchase DATABOSS/2 for use as the basis of the automated toponymic system.

DATABOSS/2

DATABOSS/2 is a conversational data-base management system written in the Basic-Plus programming language. It performs in either batch or on-line mode. DATABOSS/2 permits users to define a data base on-line and to load and manipulate the data base contents through the use of standard functions. In addition, users can include their own routines for input and output control.

All standard DATABOSS/2 functions can take input from any ASCII file. New records can be added to the data base from a disk file, a magnetic tape device, cards, paper tape or any other ASCII input device. DATABOSS/2 is currently implemented exclusively on DEC PDP mini-computers running on the RSTS/E operating system.

DATABOSS/2 was developed by Florida Computers Incorporated of Miami, a Digital Equipment Corporation (DEC) original equipment manufacturer. DATABOSS/2 has been used, or is currently being used, in such diverse applications as McDonald's Restaurants, Kentucky Fried Chicken, the Canadian Cancer Society and the federal Department of Agriculture. In the United States, users include the U.S. Air Force, Bechtel Corporation, Westinghouse, several state power utilities and the Memphis General Hospital.

DATA BASE AUTOMATION PILOT PROJECT

Once the decision had been made to undertake the automation of the toponymic data base, it was deemed essential to carry out a pilot project to determine the optimal automation process. The names records of the Northwest Territories were selected for use in a trial project. This decision was made, firstly, due to the fact that the present edition of the Northwest Territories Gazetteer had been published in 1971 as a provisional volume, and secondly, because the data base for the Northwest Territories was relatively small and was therefore conducive for use in such a project. During the period from July to September 1978, the Northwest Territories information was transcribed from existing card files onto computer coding forms. The completed forms were then sent to a word processing firm for the construction of a magnetic tape, with the tape completed in October 1978.

During that same month a contract was awarded to IAS Computers Ltd., of Halifax, to implement the toponymic data management system, by supplying the DATABOSS/2 software package and a RSTS/E operating system, and by installing this software on an existing PDP 11/45 computer held by the Topographical Survey Division, Surveys and Mapping Branch. IAS was directed to bring the system on-line in accordance with detailed storage and retrievals specifications, and to load the Northwest Territories tape onto the system. This phase was successfully completed, and in January 1979 a second contract was awarded to IAS to define and construct a data dictionary for the Northwest Territories tape, and to train Surveys and Mapping programming staff in the use of the automated system. Several computer terminals and a printer were acquired during December and January 1978-79 for use in the interactive editing of the Northwest Territories data.

Editing of the 16 800 Northwest Territories records was completed several months ago. A revised tape of the records, incorporating all changes and additions, was produced on the system and was sent to a local printing firm for production of the new volume of the Northwest Territories Gazetteer - the first such volume of the Gazetteer of Canada series to be produced from information derived from the automated toponymic data base.

This volume, using traditional printing methods, would have cost approximately $28.00 per copy to produce. Using the computer output derived from the data base as direct input into a Xerox 9700 Laser Printer, conventional typesetting costs were eliminated, allowing the gazetteer to be produced at a cost of less than $8.00 per copy. The printed gazetteer is accompanied by a set of 5 microfiche containing all the printed gazetteer information, including maps. The microfiche, in updated form, will be sent to subscribers once a year, while the hard copy book will be revised once every three to five years, depending on the frequency of information change.

The encoding of the remaining geographical names records of Canada is continuing under contract. Five operators are presently engaged in the keyboarding required to complete this project by December 1981. As well as the records of the Northwest Territories, the data for the Yukon, Manitoba and approximately half of Ontario are now on-line, bringing to nearly 90 000 (or 25% of the total), the number of names records now on the system. At the present rate of encoding, and barring any significant problems, it appears likely that the December 1981 deadline will be met.

PRESENT SYSTEM STATUS

In general, our initial conviction that a commercially available package could successfully form the basis of an automated toponymic system has been confirmed. There is little doubt that without the use of DATABOSS/2 as a foundation, the Toponymy Unit might still be waiting for a computer software house to create a comprehensive system, at perhaps far greater cost than has been incurred using DATABOSS/2.

On the other hand, the basic DATABOSS/2 package, as purchased from IAS, did not provide all the aspects considered important in the implementation of an effective names system. DATABOSS/2, for example, is fundamentally a single-file construction system, with maximum record lengths of 512 characters. While this record length works well for a large percentage of the names records, it is a constraint in the preparation of those name files which contain lengthy records. In these cases, the total number of records can easily exceed 512 characters. For this, and several other important reasons, it was decided to abandon the single-file structure of the purchased DATABOSS/2 system in favour of a multi-file system. To date, seven such files have been developed in-house for use in DATABOSS/2.

The first of these is the Gazette Data Base File. This file is divided into eleven fields of information, all relevant for gazetteer production and for general search information. It has the following fields:

(1) *Unique Key Field* - a different sequence of five letters associated with each name, e.g. AALPH-Tuktoyaktuk;

(2) *Unique Key Name Field* - the first fifteen characters of the feature name, without punctuation or special characters;

(3) *Feature Name Field* - up to 53 characters, using both upper and lower case, and any accents used in the name, with the exception of a diaresis mark (ö);

(4) *Generic Field* - a four digit numeric code used to identify the generic;

(5) *Province Field* - a two digit code used to represent the province or territory that the feature occurs in;

(6) *Decision Date Field* - a six digit numeric representation of the most recent decision date for that name, with the date ordered by year/month/day;

(7) *Status of Name Field* - a two digit alphanumeric code representing the status of the name ('A' status names are currently approved for official use, 'B' and 'C' status names are not recognized for official use);

(8) *Latitude Field* - six characters, recording the latitude in degrees, minutes and seconds;

(9) *Longitude Field* - seven characters, recording the longitude in degrees, minutes and seconds;

(10) *Obscure Generic Field* - if a 'y' is entered in this field, indicating the generic is obscure, the generic will be displayed in brackets after the name, e.g. The Mitres (hills);

(11) # *Maps for Gazetteer Field* - the number of map sheets for each name to appear in the gazetteer with a maximum of 2 map sheets per name for gazetteer purposes;

(12) *Include Name Always Field* - gazetteers contain all the officially recognized names; if a name is no longer officially recognized but, for various reasons, is to appear in the gazetteer a 'Y' is entered in this field; and

(13) *Expansion Field* - 25 characters to allow for future changes to the length of existing fields or the creation of new fields.

The second file is a Map Sheet File, containing the numbers of all map sheets and charts on which a feature is located. It is divided into three fields of information:

(1) *Unique Key Field* - this key corresponds to the Unique Key in the Gazette Data Base File;

(2) A five character *Sequence # Field;* and

(3) *Map Sheet or Chart # Field* - a six digit alphanumeric representation of the map sheet or chart on which a feature is located.

The other five data base files have similar structures. Files number three and four are Location Data Base Files containing cadastral information. The fifth file is the Cross Reference Data Base File, the sixth is the Location Narrative Data Base File, and the final file is the Origin Narrative Data Base File.

Three programs have been developed in-house to facilitate data inputting, editing and retrieval on the DATABOSS/2 system.

The first of these is the Interactive Data Entry program. This program creates temporary data files that are filled by contract operators as they interactively keyboard their data. File outputs are checked daily by a supervisor before being transferred, by authorized staff, to permanent data base files. The Interactive Data Entry program guides the terminal operator through each of the fields of information required to construct a proper data record. It asks for "Name?", "Decision Date?", etc., and will not proceed to the next prompt until either a proper response (i.e., correct combination of letters and/or numbers) or a default ("no-data" signal) is given. At the end of the session, the operator reviews and corrects the data and obtains a hard copy of the CRT display from the printer. The printout is passed to a supervisor for audit, and the operator then begins the construction of the next temporary file.

The second program is the Bulk Entry Program. This program provides the supervisor or authorized editor with the ability to merge temporary, operator-created Interactive Entry program files with the permanent data base. The Bulk Entry program massages the temporary file information into a format acceptable to DATABOSS/2. This manipulation includes the reduction of the inputting information into its seven DATABOSS/2 constituent file parts, and the integration of those parts into their appropriate data base records files.

The final program is a Report Generator, which combines, in a user transparent process, all related names data requested for a specific name during a system search. Without the use of such a program, operators on the system would be required to go to each of the various files on which information pertaining to a specific name is held and search those files individually, later manually combining all search outputs in order to consolidate the required information for a given name.

The following operations, on all or selected fields, can now be carried out on the automated system:

(1) create a new data base file;

(2) display the data base file layouts;

(3) enter a new record into the data base;

(4) remove a record from the data base;

(5) display a data base record;

(6) change a non-key field in a record; and

(7) change any field in the data base.

Security

At the present time, the system is relatively secure against inadvertent and deliberate damage.

Information areas (files and records) cannot be accessed without the use of controlled passwords, and records cannot be changed or deleted without a knowledge of the required protection codes. All retrieval steps are protected by passwords and the Unique Key field and the Unique Key Name field are locked to all unauthorized users.

Daily tape backups are made from current disks, and are stored on tape in a secure site. Arrangements have been completed for a "buddy-system" backup whereby an outside computer bureau also using DATABOSS/2 is engaged in a system partnership that allows either participant to share the other's system, should a catastrophic failure occur at either site.

SYSTEM HARDWARE

Various hardware components have been acquired during the development of the system. Currently, the system is comprised of the following items:

(1) 1 x DEC PDP 11/45 computer;

(2) 1 x 128K words memory;

(3) 1 x 1 TU-10 800 bpi tape drive;

(4) 4 x RJPO4 disk drives, each 88 Megabytes;

(5) 1 x DZ-11 multiplexor - 8 lines;

(6) 1 x DC-11 communications interface (dial up);

(7) 1 x HP 2631A printer;

(8) 4 x Volker-Craig 404 terminals with modems;

(9) 1 x Volker-Craig 414 Word Processing terminal with modems; and

(10) 1 x Cybernex CRT.

SYSTEM COST

To date, a direct investment of approximately $325 000 has been made in the automated system, including salaries, operational funds and capital, but exclusive of an existing computer hardware investment of nearly $151 000. Total system value is currently $475 000. It is estimated that, over the next two years, an additional $185 000 will be required to fund the volume encoding of those records not available in machine-readable form. Another $50 000 will be spent on the conversion of machine-readable information from the provinces of Quebec, Manitoba, Alberta and Ontario. Anticipated total system investment by the December 1981 planned completion date is approximately $700 000, with a recurring annual software and hardware maintenance budget of $25 000. The single largest cost component in the system automation implementation is being incurred during the encoding of the large volume of available data.

The overall expenditure committed to the development of the data base achieves its proper perspective when it is remembered that system implementation will allow the 12 members of the Toponymy Unit staff to now work in concert with the members of the CPCGN Secretariat in keeping abreast of toponymic requirements, without additional human resource requirements. Indeed, if automation were not implemented, external reviews indicate a requirement by the end of the current fiscal year, of 22 person-years in the toponymic activity in order to meet ever-increasing geographical name data processing and information demands.

Expenditure forecasts for this 22 person-year, conventional-type data base predict annual costs of $526 000 by 1983-84. In comparison, the 12 person-year, automated system cost estimates indicate an average annual cost, by 1983-84, of $359 000 - an annual saving of approximately $167 000. In addition, it is anticipated that with the adoption of the automated system, two clerical positions can be converted to professional positions by December 1981, to begin to improve the overall quality of the data base through the administration of field research contracts coordinated to federal and provincial mapping requirements.

CONCLUSIONS

I would make the following conclusions regarding the Toponymy Unit's on-going experience with the automation of the National Toponymic Data Base:

(1) Prepare, and understand thoroughly, a detailed problem specification, quantifying as precisely as possible your present and future system requirements and expectations.

(2) Have faith in your problem specification once it has been developed, and use it continually to measure system progress. Modify the specification only if there is no other reasonable alternative.

(3) Do not accept as gospel, the advice of consultants, contractors or computer software houses. Their objectives, however honourable, may not be your objectives. Continual vigilence is required to ensure that every party involved in the system development project shares the same understanding of the problem specification, and that everyone is working towards a well documented common objective.

(4) Do not accept more, or less, than your problem statement calls for. Beware of elaborate, custom-designed systems that in the long run do nothing more than increase your unit cost.

(5) Rent, rather than buy, hardware if at all possible. Rapidly changing technology can be used more effectively to your advantage if purchases are kept to a minimum.

(6) Finally, though it may sound trite, the familiar expression: "Make your best time and cost estimate, double it and you will only be short by 50%" was, I feel, intended specifically for the development of an automated names system. Do not underestimate the amount of time and effort required for such an undertaking.

SYSTÈME AUTOMATISÉ DE GESTION DE DONNÉES TOPONYMIQUES

Pierre Barabé*

RÉSUMÉ. La Commission de toponymie estime nécessaire de mettre en place le système ØNØMA compte tenu de l'écart entre, d'une part, le volume d'opérations qui découle de l'exercice des juridictions nouvelles (inventaire, normalisation, officialisation et diffusion) que la *Charte de la langue française* lui attribue et d'autre part, les possibilités du système actuellement utilisé. La Commission se situe au centre du système de diffusion de données toponymiques. Si le projet ØNØMA se réalise, nous favoriserons une plus grande accessibilité aux données, en assurerons le contrôle et réduirons les risques d'altération de l'information lors de la diffusion. Des mécanismes permettant la compatibilité avec d'autres systèmes de gestion seront étudiés afin de faciliter l'échange d'informations toponymiques.

ABSTRACT. Given the increased volume of operations in light of the new functions (inventory, standardization and gazetting) it must assume, in accordance with the *Charte de la langue française,* and the limitations of the system presently in use, the Commission de toponymie du Québec has deemed it necessary to implement the ØNØMA system. The Commission de toponymie is at the centre of the toponymic data distribution system. Were the ØNØMA system in use, accessibility to data would be increased, its control ensured and the risk of alteration of the distributed information reduced. Mechanisms to ensure the compatibility of ØNØMA with other management systems will be studied in order to facilitate the exchange of toponymic information.

Le Mémoire sur la conception administrative d'un système automatisé d'une banque de données toponymiques, identifié comme étant le système ØNØMA fait suite à une révision des moyens et des effectifs dont dispose la Commission de toponymie afin de répondre adéquatement aux pouvoirs et devoirs nouveaux déterminés par le mandat qui lui a été confié par la *Charte de la langue française* le 26 août 1977.

Après deux ans et demie de fonctionnement en vertu de la loi qui l'a créée et selon la nouvelle structure dont elle s'est dotée, la Commission de toponymie estime nécessaire de mettre sur pied un nouveau système, compte tenu de l'écard entre, d'une part, le volume d'opérations qui découle de l'exercice des juridictions que la loi lui attribue et, d'autre part, les possibilités du système actuellement utilisé.

JURIDICTIONS DE LA COMMISSION DE TOPONYMIE

La *Charte de la langue française* a en effet confié à la Commission de toponymie des juridictions considérablement élargies par rapport à celles qui caractérisaient l'ancienne Commission de géographie. Celle-ci n'avait, somme toute, qu'une juridiction de contrôle sur la toponymie apparaissant sur la cartographie, en plus d'un certain pouvoir d'édicter des règles d'écriture des noms de lieux.

La nouvelle loi confère à la Commission de toponymie une juridiction plus large quant aux types de noms de lieux concernés; la loi n'apporte aucune limitation quant à la définition de noms de lieux, de sorte que compte tenu cependant des juridictions concurrentes avec certains organismes, notamment en matière municipale, la Commission a juridiction, en

* Pierre Barabé, Commission de toponymie du Québec.

dernier ressort, sur tout type de nom de lieu au Québec. Par ailleurs, les juridictions de la Commission sont également amplifiées, si l'on peut dire sur le plan vertical, conférant aux décisions de la Commission de toponymie un caractère obligatoire. Aussi, la Commission s'est vue confier un pouvoir réglementaire sur les critères de choix, les règles d'écriture et la procédure, pouvoir que n'avait pas l'ancienne Commission. Enfin, les moyens mis à la disposition de la Commission de toponymie par le Gouvernement du Québec témoignent de la volonté de vouloir élargir l'exercice des juridictions: notamment, les ressources humaines de la Commission ont été plus que doublées depuis 1977.

L'élargissement des juridictions de la Commission gravitant autour de quatre champs principaux, soit l'inventaire, la normalisation, l'officialisation et la diffusion, se traduit:

(1) par la diversification de ses opérations qui s'insèrent dans un contexte toponymique quantitativement important;

(2) par l'augmentation des débits d'entrée, de traitement et de sortie; et

(3) par l'augmentation de la documentation.

En effet, environ 120 000 toponymes sont actuellement consignés à la Commission dont 81 000 sont officialisés. Nous estimons également qu'il existerait environ 125 000 toponymes (incluant les banques de données odonymiques) consignés ailleurs qu'à la Commission, qu'il y aurait plus de 200 000 noms en usage et qu'environ 1 500 000 entités seraient encore innommées, d'où la nécessité de systématiser les opérations constituant les différentes étapes du processus de désignation, tout particulièrement celle de l'officialisation, de même que le cheminement de l'information.

SYSTÈME ACTUEL

L'ancienne Commission de géographie s'était dotée, en 1971, d'un système d'exploitation par ordinateur. Ce dernier avait été établi afin de conserver un fichier à jour de tous les noms de lieux officiels du Québec. On a pu de la sorte vérifier systématiquement les noms officiels en vue de la réédition et de la refonte du *Répertoire toponymique du Québec* (75 000 noms) et de la publication périodique à la *Gazette officielle du Québec* des noms approuvés par la Commission. Ce système, encore en usage aujourd'hui, est axé uniquement sur la vérification, la mise à jour et la diffusion des noms de lieux officiels. Il permet également la sortie de listes sélectives, nécessaires au travail de la Commission.

En faisant reposer son travail sur une banque de données composée à la fois d'un fichier manuel et mécanographique, la Commission est aux prises avec un problème réel d'accessibilité aux données, qui débouche bien souvent sur un égarement temporaire de l'information à cause du volume de fiches de plus en plus important (au-delà de 120 000) et du nombre de personnes qui ont à les manipuler pour en faire la mise à jour.

De plus, le statisme du système mécanographique actuel, ne faisant état que du caractère officiel de la toponymie québécoise, ne nous permet pas de suivre l'évolution des toponymes, aussi bien dans le temps que dans le processus d'officialisation.

Si bien que le système de gestion de données toponymiques présentement en opération à la Commission de toponymie ne satisfait qu'en partie seulement aux objectifs que poursuit celle-ci pour réaliser les devoirs et pouvoirs dévolus par la loi.

NOUVEAUX UTILISATEURS

Comme notre organisme a maintenant une juridiction de "dernier ressort" sur tous les types de noms de lieux, il doit répondre à de nouveaux utilisateurs qui jusque-là n'étaient pas tenus de nous consulter. Actuellement ils se composent presqu'exclusivement d'organismes véhiculant des noms de voies de communication. Ainsi à moyen terme, nous aurons à répondre aux exigences des organismes suivants qui veulent se conformer à la loi: le ministère des Transports, les municipalités, les commissions scolaires, Hydro-Québec, la Régie des services publics et le Bureau du directeur général des élections. À plus long terme, notre action se prolongera au niveau des organismes privés qui ont à rejoindre le citoyen québécois, par exemple les entreprises de télécommunications.

Comme on estime à plus de 100 000 le nombre d'odonymes ruraux et urbains sur lesquels la Commission doit exercer un contrôle, il devient donc nécessaire de multiplier les composantes du système et d'introduire de nouvelles variables pour pouvoir répondre, dans un délai raisonnable, aux demandes variées de ces nouveaux utilisateurs.

OBJECTIFS GÉNÉRAUX DU SYSTÈME ØNØMA

La mise en place du système ØNØMA permettrait l'atteinte d'objectifs essentiels à l'efficacité de la Commission de toponymie:

(1) Une production croissante reliée à une augmentation sensible de la demande et des besoins en matière de toponymie.

Cet objectif concerne particulièrement deux des devoirs de la Commission qui sont la conservation des noms de lieux et leur officialisation.

(2) Une réduction des délais d'exécution et d'attente reliée à la diligence et à la célérité avec lesquelles doit agir la Commission, particulièrement parce qu'elle joue un rôle de service et, à ce titre, ne constitue, bien souvent, qu'un des chaînons d'une activité à incidence économique importante.

Cet aspect touche tant les processus d'analyse, de l'officialisation et de la diffusion des noms de lieux que ceux enclenchés lorsqu'il s'agit de répondre à des consultations ponctuelles.

(3) Un traitement toponymique de qualité et un produit exact:

L'objectif relié au premier volet vise la création, la normalisation et l'analyse des noms de lieux alors que celui relié au second volet touche la diffusion et l'utilisation d'une toponymie exacte.

En résumé, une centralisation toponymique, résultant de la mise en place du système ØNØMA, devrait favoriser l'exactitude du résultat des recherches et faciliter, en particulier, le respect du principe fondamental en toponymie qui veut qu'un lieu ne soit désigné que par un seul nom.

DESCRIPTION ET CARACTÉRISTIQUES DU SYSTÈME ØNØMA

Avant d'aborder ce chapitre, nous tenons à vous préciser qu'il s'agit d'un projet et que la réalisation pourra être sensiblement différente de ce que l'on projette. En effet, même si le Conseil du trésor a accepté en principe notre mémoire, il reste que, avant de prendre une décision finale sur l'implantation et le développement du nouveau système, la Commission de toponymie devra fournir certaines précisions premièrement quant aux mécanismes qui devront être mis sur pied pour permettre l'intégration des différents corpus toponymiques déjà existants, deuxièmement quant aux effectifs qui seront nécessaires pour effectuer l'intégration des données contenues dans notre propre banque et dans les banques extérieures et finalement quant aux coûts/bénéfices du logiciel à utiliser.

Les fichiers

Le système ØNØMA sera constitué d'un fichier principal, le fichier toponymique et de trois fichiers secondaires:

(1) le fichier "noms en réserves" qui permettra l'accélération des recherches pour les fins de désignation;

(2) le fichier "sources" ayant pour objet de donner accès aux ouvrages d'où origine l'information concernant la date de la première attestation du nom, sa signification et son étymologie; et

(3) le fichier "chercheurs" permettant d'identifier la personne qui est l'auteur d'une partie de la fiche toponymique, particulièrement pour les informations concernant la signification, l'étymologie et la prononciation.

Dans le fichier toponymique, on réfèrera aux fichiers "sources" et "chercheurs" au moyen d'un code correspondant à la classification des sources documentaires dans le cas du premier fichier et au numéro du chercheur dans le cas du deuxième.

Nous nous attarderons plus particulièrement au fichier toponymique, le plus important à cause du volume d'informations qu'il contient.

Même si sa conception n'est pas définitive, la fiche toponymique devant servir à consigner l'information, contiendra les rubriques suivantes:

(1) L'identification du toponyme faisant état de son statut et de son genre. On a prévu la codification du générique, des particules de liaison et de l'entité.

(2) Les sources d'attestation du toponyme au moyen d'un code.

(3) Le numéro de codification des toponymes utilisé par la banque de données extérieures permettant ainsi l'échange d'informations.

(4) L'état de l'officialisation comprenant la date de même que la codification des critères qui servent à l'officialisation ou à la non-officialisation des toponymes.

(5) La localisation par coordonnées géographiques (pour les odonymes, nous retenons les coordonnées des aboutissants et tenants en ayant comme principe que les premières seront toujours celles du point situé le plus à l'est), par code géographique, par carte topographique (1/50 000 et 1/20 000), par carte marine, par division de recensement, par canton et par région administrative.

(6) L'échelle de parution.

(7) Les données linguistiques et historiques: code linguistique, code sémantique, prononciation, chercheur, première attestation connue.

(8) L'origine: étymologie, motifs, signification, sources.

(9) Le renvoi au nom officiel d'une même entité ou d'une entité substituée.

(10) Les autres renseignements que nous ne pouvons pas consigner dans une ou l'autre des rubriques mentionnées précédemment.

Intrants du système ØNØMA

En procédant à la conservation des noms de lieux, la Commission de toponymie doit voir, premièrement, à constituer sa propre banque de noms géographiques; deuxièmement, à centraliser l'information toponymique existant ailleurs en vue de favoriser une plus grande accessibilité des données normalisées en fonction de ses normes et critères. Un système intégré d'information, de référence et de transfert permettra de minimiser les risques de perte ou d'altération de cette information.

Les intrants devant constituer le système préconisé peuvent être groupés sous deux titres:

(1) Les intrants originant de la Commission de toponymie, c'est-à-dire toute l'information toponymique recueillie depuis la naissance de l'ex-Commission de géographie jusqu'à nos jours. Cette documentation se présente sous divers aspects: fichier mécanographique, fichier manuel, cartes anciennes, dossiers, etc.

(2) Les intrants originant de l'extérieur. Ils comprennent:

- le fichier des odonymes du Bureau du directeur général des élections (65 000 odonymes);
- le fichier des toponymes populaires du Bureau de la statistique du Québec (10 500 toponymes);
- le fichier des plans d'eau du ministère de l'Énergie et des Ressources (60 000 plans d'eau);
- le fichier des cours d'eau du ministère de l'Énergie et des Ressources (6 000 cours d'eau);
- le code géographique du Bureau de la statistique du Québec (identifiant chacune des municipalités);
- les découpages territoriaux; et
- le fichier des odonymes du Centre de recherche en aménagement régional de l'université de Sherbrooke (20 000 odonymes).

Fonctions du système ØNØMA

Les fonctions dévolues au système regroupent les différentes activités réalisées par les Services de la Commission. Ces activités, qui supportent et alimentent la fonction d'officialisation assurée par les Commissaires, apparaissent comme des éléments du processus toponymique.

Sept fonctions ont pu ainsi être déterminées: alimentation, traitement toponymique, création et mise à jour des banques, officialisation, diffusion, contrôle et promotion. Pour chacun comme pour l'ensemble des noms de lieux du Québec, les fonctions du système ØNØMA s'insèrent dans un cycle que l'on pourrait caractériser de la façon suivante: "de l'usage spontané à l'usage contrôlé".

Les extrants

L'ensemble des opérations de la Commission de toponymie s'oriente vers la production de matériel toponymique sous diverses formes (listes, répertoires, cartes, tableaux statistiques) destiné au travail interne des services de la Commission, mais ultimement aux utilisateurs à qui notre organisme a le devoir de diffuser la nomenclature géographique.

La spécialité des besoins des utilisateurs gouvernementaux et celle des opérations de traitement que les services de la Commission doivent réaliser pour rencontrer à la fois les besoins et les objectifs que la Commission s'est fixés exigent de dégager avec précision les extrants que le système doit générer.

Voici les principaux types d'extrants requis par les besoins spécifiques des utilisateurs:

(1) publication annuelle des noms approuvés par la Commission à la Gazette officielle;

(2) Répertoire toponymique du Québec (tous les 5 à 6 ans);

(3) listes sélectives comprenant:

- répertoires sectoriels (lieux habités, odonymes, plans d'eau, cours d'eau, réserves indiennes et établissements autochtones);
- listes par feuillet cartographique à différentes échelles;
- listes par découpages administratifs;
- listes par génériques ou par entités;
- listes par status de nom;
- listes par langues;
- listes par caractéristiques linguistiques;
- listes d'homonymes;
- listes thématiques (hagionymes, anthroponymes, toponymes topographiques, ...).

(4) consultation directe;

(5) statistiques; et

(6) listes de contrôle.

Certaines sorties peuvent contenir des éléments jugés non nécessaires au niveau des intrants. Pour éviter la lourdeur du système, il est apparu opportun de prévoir que certaines données existant dans d'autres systèmes puissent demeurer extérieures au système ØNØMA, tout en prévoyant leur intégration aux schémas de sortie pour la compatibilisation des programmes.

Il y a lieu de préciser la distinction entre "listes" et "répertoires": les premières sont essentiellement des instruments de travail pour les organismes de l'Administration et les seconds constituent un produit plus raffiné, destiné à une plus large diffusion. Des programmations similaires seront utilisées et permettront d'automatiser la production de répertoires.

Particularités du système

Compte tenu des objectifs visés par la Commission et des types de besoins exprimés par les différents utilisateurs, il apparaît essentiel de prévoir certains éléments particuliers dans la description du système à mettre sur pied.

Ces prévisions touchent:

(1) Les modes d'accès à l'information. Ainsi nous privilégions l'installation d'un système d'accès direct au fichier-maître au moyen d'un écran cathodique.

(2) L'utilisation de la chaîne française.

(3) Le traitement particulier des toponymes amérindiens et inuits exigent l'utilisation de signes diacritiques. L'absence d'une normalisation absolue dans la représentation écrite de cette catégorie de toponymes nous oblige pour l'instant à n'utiliser qu'un indicateur révélant la présence de caractères spéciaux inscrits sur ces noms.

(4) Les modes de restitution de l'information à diffuser sous forme de listes, de répertoires (y compris la Gazette officielle) et de cartes.

(5) Certaines programmations particulières comme par exemple la possibilité d'obtenir des listes, par ordre alphabétique, des toponymes présentés dans leur forme inversée. Ces listes permettront en l'occurrence l'étude des racines suffixales des toponymes amérindiens ou inuit, une question reliée à la normalisation de ces toponymes.

IMPACT D'IMPLANTATION

Sur les utilisateurs

Du fait de la centralisation entre les mains d'un seul organisme de toutes les données toponymiques et la possibilité pour chaque utilisateur d'accéder rapidement à cette banque d'informations afin de satisfaire ses besoins spécifiques, résulteront des économies appréciables de temps-travail par le fait même une meilleure utilisation des effectifs.

Ainsi, certains organismes et ministères à vocation territoriale éviteront de mobiliser des effectifs afin d'effectuer des relevés de noms sur le terrain pour satisfaire des besoins spécifiques ou urgents nécessaires à la justification et à la réalisation d'objectifs ou de politiques parce que, avec la mise en place du système ØNØMA, la Commission pourra leur fournir dans les délais prévus l'information désirée.

En second lieu, certains organismes préoccupés par l'identification de découpages administratifs ou de lieux ponctuels ou linéaires éviteront de consacrer souvent plusieurs jours à des recherches axés sur la pertinence ou la justification d'un nom susceptible d'identifier un lieu. En effet, le système ØNØMA retient le principe de la constitution d'une réserve de noms potentiellement utilisables pour les diverses catégories de lieux à nommer.

En dernier lieu, les organismes concepteurs de cartes verront leur production augmentée par ce système qui permettra une réduction du processus de vérification et de mise à jour de la toponymie sur les documents cartographiques tout en assurant une amélioration qualitative du document par l'inscription exacte et au bon endroit des noms de lieux. C'est là un objectif que tend à atteindre tout organisme préoccupé de rentabilité administrative.

Sur la Commission de toponymie

L'entrée dans le système ØNØMA de données d'ordre historique rattachées aux toponymes, permettra d'une part de fournir dans la majorité des cas la réponse à des consultations en évitant aux chercheurs la tâche de référer à une multitude de sources documentaires ou archivistiques éparpillées en maints endroits.

Du point de vue terminologique, le nouveau système permettra aux chercheurs de cerner plus rapidement entre autres les contextes d'utilisation des termes génériques et des entités, d'obtenir en des délais réduits des représentations statistiques des phénomènes étudiés et de pouvoir rendre promptement les résultats de leurs analyses.

De même ce système permettra-t-il d'éviter tout dédoublement de travail inutile et coûteux. En effet, en fournissant sur demande les listes à jour de toponymes extraits du système, la Commission n'aura plus à inscrire elle-même sur les documents présentés par les utilisateurs les toponymes officiels, opération qui nécessite actuellement beaucoup de temps.

Et enfin parmi les résultats positifs de l'instauration du système ØNØMA sur les activités de la Commission, notons également le fait que celui-ci favorisera une plus grande capacité d'analyse des dossiers relatifs à l'officialisation des toponymes ou au contrôle de leur qualité.

Sur les systèmes existants

La Commission de toponymie étant l'organisme officiel par où doivent passer tous les utilisateurs de données toponymiques, aussi bien du Québec que de l'extérieur, nous favorisons avec ce système une plus grande accessibilité aux données, en assurons le contrôle et réduisons les risques d'altération de l'information lors de la diffusion.

La Commission de toponymie n'a pas l'intention de mettre en place un système inaccessible aux utilisateurs ayant déjà un système mécanographique en usage. Elle prévoit donc élaborer des mécanismes permettant la compatibilité avec d'autres systèmes de gestion afin de faciliter les échanges d'informations et répondre ainsi à des besoins de rentabilité administrative.

Dans un premier temps, nous élaborons des protocoles qui préciseront les mécanismes de communication avec des organismes opérant au Québec.

Dans un deuxième temps, nous faciliterons l'accès à l'information toponymique à tout autre organisme opérant à l'extérieur du Québec, entre autre au Comité permanent canadien des noms géographiques afin de répondre à l'ensemble des problèmes de toponymie au Canada.

TRAITEMENT DES NOMS GÉOGRAPHIQUES À L'AIDE DE SYSTÈMES D'INFORMATION AUTOMATISÉS À RÉFÉRENCE SPATIALE

C.B. Berthold* et B.J. Rivard**

RÉSUMÉ. L'un des objectifs de cet exposé est de présenter très brièvement, le concept de "système d'information automatisé à référence spatiale", à travers l'une de ses principales composantes à savoir la "base géographique", afin d'analyser l'incidence particulière et notoire des noms géographiques en termes d'entrée et d'emmagasinement, de même que des possibilités de traitement et de sortie par l'apport de la dimension spatiale et de la représentation graphique automatisée.

* C.B. Berthold, Analyste-programmeur, Centre de recherche en aménagement régional, Université de Sherbrooke.

** B.J. Rivard, Directeur adjoint, Centre de recherche en aménagement régional, Université de Sherbrooke.

ABSTRACT. One of the objectives of this paper is briefly to present the concept of a "spatial referencing automated information system", notably through one of its main components, the "geographical base". Outlined is the spatial treatment of geographical names, relative to their input, storage, and processing and output possibilities, through the use of the spatial dimension and the computer generated plot.

INTRODUCTION

Nous sommes reconnaissants envers les représentants du Comité permanent canadien des noms géographiques, de même qu'à ceux de la Commission de toponymie du Québec, de nous permettre d'exposer brièvement, à ce colloque, l'une des nombreuses facettes que comportent "L'automatisation des noms géographiques", notamment celle de leur traitement au moyen de systèmes d'information automatisés à référence spatiale.

Nous avons été sensibilisés et appelés à considérer l'incidence importante que pouvait exercer la toponymie, non pas en tant que spécialistes des noms de lieux géographiques en matière de définition, d'officialisation ou d'approbation, ni même en tant qu'agents créateur ou d'agents de contrôle, mais bien en tant que spécialistes de la transposition (géocodification), en ordinateur des différentes configurations spatiales (ex: réseaux routier, ferroviaire, hydrographique, lotissement, assises d'aménagement, limites d'intérêt, etc.), pouvant composer un territoire donné, afin de pouvoir établir les relations "données-espace", et permettre leur traitement et leur représentation graphique automatisée en conformité avec le lieu auquel elles se rattachent pour les fins appliqués d'analyse, de simulation et de gestion territoriale, tant au niveau municipal, urbain que régional, pour le compte de divers organismes publics, parapublics et privés, oeuvrant à divers paliers d'intervention et dans des domaines d'activités les plus variés.

SYSTÈME D'INFORMATION AUTOMATISÉ À RÉFÉRENCE SPATIALE

Ce type de système qui se caractérise par le concept de "référence spatiale", se distingue des systèmes de traitement dits conventionnels, de par la présence de la "base géographique" constituant l'une de ces trois principales composantes en ordinateur.

Cette base géographique comporte un ensemble de fichiers autonomes mais interreliés où sont emmagasinées les données spatiales en termes de pointeurs géographiques (identifiant: nomenclatures et codes des éléments spatiaux géocodifiés; repérant: numéros des noeuds et centroïdes; et localisant: coordonnées X, Y/UTM, MTM, géodésiques, etc...) et informatiques, permettant de transposer et de simuler en ordinateur les configurations des éléments choisis pour représenter la réalité spatiale d'un territoire donné.

De plus, cette base géographique est mise en relation, à l'aide d'un module de traitement comportant le logiciel (programmes et sous programmes) permettant sa gestion et son entretien (création, mise-à-jour, sortie de listes et de représentations graphiques automatisées), de même que son exploitation (traitement de données/applications diverses) par l'établissement des relations "données-espace", avec la base de données qui peut se composer de divers fichiers de données pertinentes (statistiques, physiques, etc...), pouvant provenir de sources diverses (ex.: Statistique Canada, bureau de la Statistique du Québec, organismes para-publics, privés, etc...) et de nature diverse, constituée selon les besoins spécifiques de chaque utilisateur, permettant le traitement et la représentation graphique des données, en conformité spatiale avec le lieu où se déroule l'activité analysée (re: Figure 1).

Depuis bientôt dix ans qu'il oeuvre dans ce domaine, le Centre de recherche en aménagement régional (CRAR) a conceptualisé, développé, opérationnalisé, et implanté un certain nombre de ces systèmes de gestion appliqués à divers niveaux d'intervention, niveau régional (re: système Info-Estrie et système APOGÉE), niveau municipal (re: système d'information Saint-Jérôme) et notamment, dernièrement (79-80), en collaboration avec les commissions scolaires du Québec (DICOS[1]) et le SIMEQ[2], la fonction gestion et exploitation de la base géographique du système collectif "GALILÉE", qui se définit comme "un système d'information automatisé à référence spatiale pour fin de gestion scolaire au Québec" système qui, de par la nature des applications demandées, en outre pour les fins de transport scolaire, fait appel à la standardisation et à la normalisation des adresses de la clientèle scolaire au Québec, d'où le concept de "toponyme" prend toute son ampleur avec les dimensions des catégories dites "officiel", "populaire", "alias" ou "parallèle", distinction importante dont nous discuterons ultérieurement à la présentation sommaire des fichiers de la base géographique, plus spécifiquement ceux dits de "Nomenclature" reflétant l'incidence de la toponymie.

Photo: Mike Kelly, Information EMR

SYMPOSIUM SPEAKERS/ORATEURS DU COLLOQUE

(Left to right/de gauche à droite): Bernard Rivard, Claire Berthold.

---

1 DICOS: Développement informatique des commissions scolaires.

2 SIMEQ: Service informatique du ministère de l'éducation du Québec.

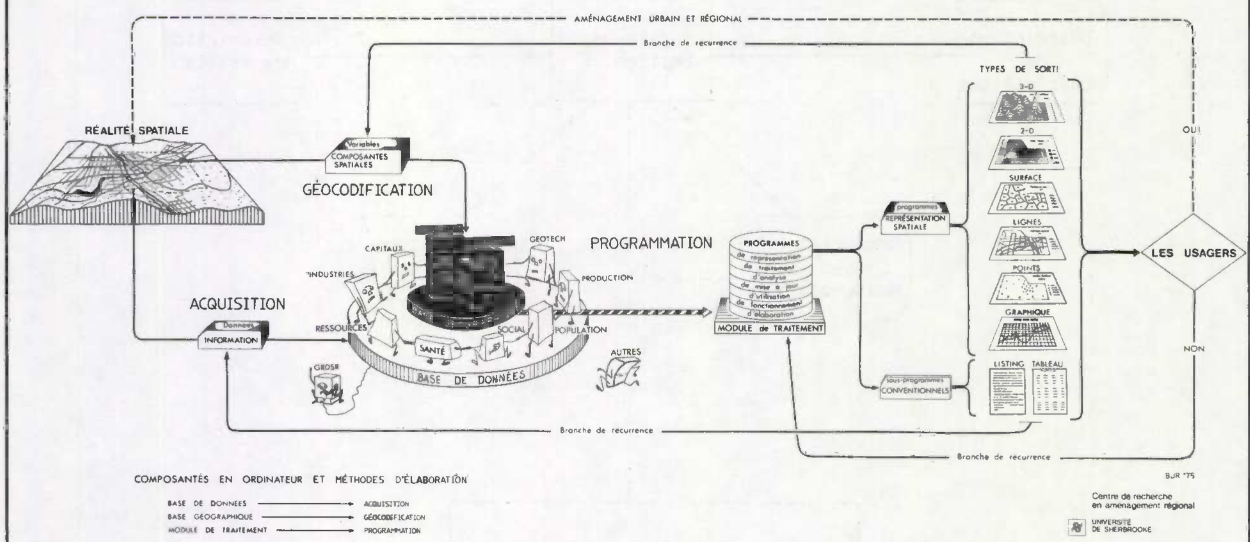

Figure 1 Schéma théorique d'un système d'information automatisé à référence spatiale

BASE GÉOGRAPHIQUE/SYSTÈME GALILÉE

La base géographique représente la dimension spatiale du système GALILÉE. Par ses six fichiers, où l'on retrouve, regroupés selon différents niveaux d'interrelations, un ensemble de pointeurs géographiques (codes de rues, segments, limites...(identifiant), numéros de noeuds et/ou centroïdes (repérant), coordonnées X et Y/UTM (localisant) et de pointeurs informatiques, on peut arriver à simuler en ordinateur, et à représenter, au moyen d'imprimante analogique (table traçante), la réalité spatiale (réseau routier, limites, etc...), du territoire d'une commission scolaire donnée. On peut également, par ces mêmes pointeurs, faciliter l'établissement des relations données-espace, lors du traitement de données diverses (clientèle, bâtisses, transport, etc...).

Les six fichiers (re: Figure 2), composant la base géographique sont: le fichier du "Découpage cartographique", qui regroupe les informations concernant les cartes couvrant le territoire ainsi que leur échelle respective, le fichier des "Noeuds et centroïdes", qui contient les coordonnées X et Y/UTM de chacun des points utilisés dans la base géographique, accompagnées d'un numéro unique assigné automatiquement à chaque point; les fichiers de nomenclature, en l'occurrence celui de la "Nomenclature des municipalités et des limites d'intérêt", contenant les noms et codes officiels de chacune des municipalités touchées par une commission scolaire, ainsi que les noms et codes d'autres limites déterminées par chaque organisme, pour ses propres besoins. L'autre fichier de nomenclature est celui de la "Nomenclature des traits du réseau routier", où l'on retrouve, groupés par municipalité, les noms

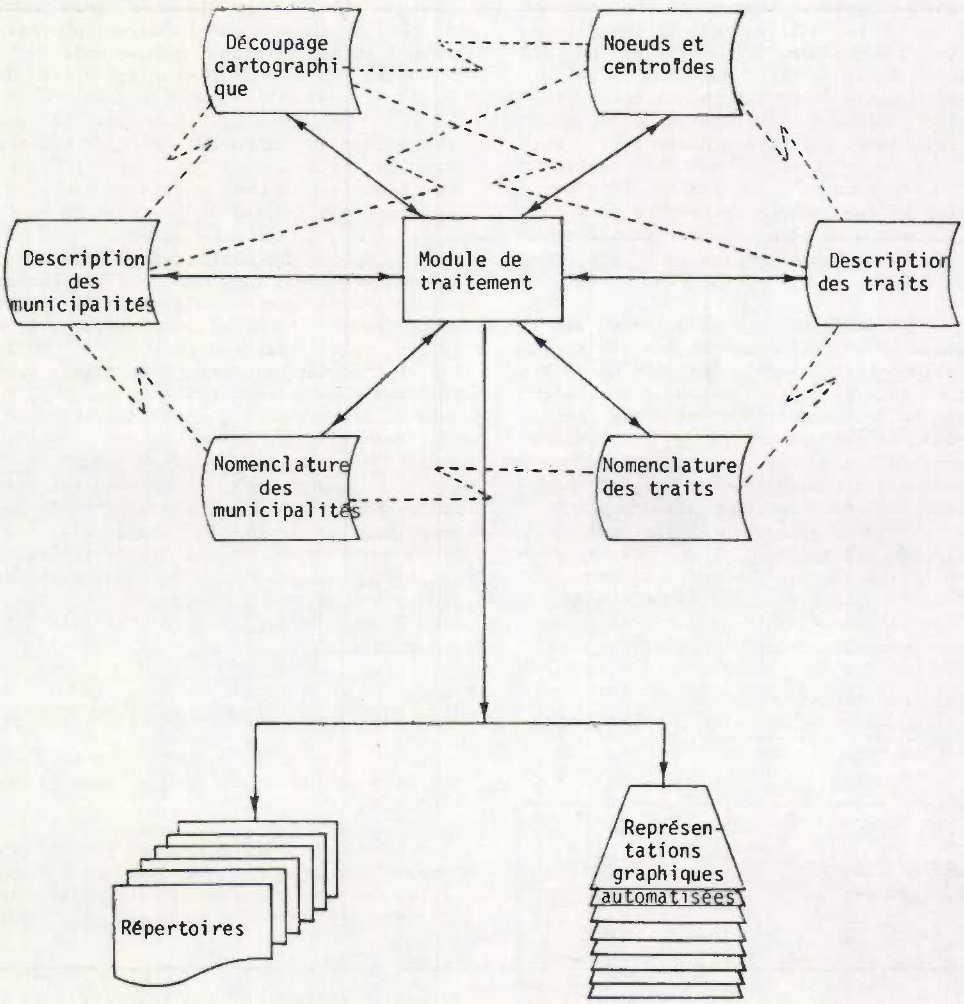

Figure 2 Schéma de la base géographique/système GALILÉE

de tous les traits du réseau routier, ainsi que les diverses toponymies populaires ou parallèles particulières à chaque trait. Un code de trait unique est automatiquement assigné à chaque trait du réseau routier différent, et il est aussi inclus dans ce fichier. Ce code sert, entre autres, à faire le lien avec le fichier suivant: "Description des traits du réseau routier".

Ce fichier contient la description, en termes de numéros de noeuds, de chacun des traits du réseau routier. C'est ce fichier, lié au autres fichiers de la base géographique, qui servira à reproduire, soit en ordinateur, soit par représentation graphique automatisée, la configuration de la réalité spatiale auxquels divers éléments pourront être rattachés. Le dernier fichier, "Description des municipalités et des limites d'intérêt", est semblable au précédent, mais représente plutôt la configuration des limites municipales, ou des limites dites "d'intérêt".

TRAITEMENT AUTOMATISÉ DES NOMS GÉOGRAPHIQUES

Avant d'en arriver à un fichier de nomenclature dont les informations sont utilisables ou traitables par informatique, plusieurs étapes de travail manuel doivent être franchies. La première étape consiste à relever tous les noms de municipalités et de limites d'intérêt pour le territoire concerné, en relevant à la fois les toponymies officielles, lorsqu'elles existent, et les toponymies populaires ou parallèles, qui peuvent aller d'une simple variante orthographique jusqu'à un nom totalement différent. Cette étape est relativement facile, grâce aux documents publiés sur la question des noms et codes officiels de municipalités par divers ministères. Mais l'étape suivante, en l'occurence le relevé des noms de traits (rues, routes, etc...) du territoire à couvrir est tout autrement, nécessitant souvent des longues recherches pour en arriver à trouver des documents et plans de base dont les informations relatives à la toponymie des lieux sont très souvent contradictoires.

À partir de ces relevés, l'étape suivante en est une de normalisation ou de standardisation des noms de traits, où l'on transcrit manuellement et selon un ensemble de règles précises, chacun des noms relevés à l'étape précédente, afin de pouvoir l'emmagasiner en ordinateur sous une forme facilement manipulable ou traitable par des programmes appropriées. Ces programmes produiront, selon le cas, des répertoires de la nomenclature sous sa forme d'emmagasinement ou sous une forme éditée, des représentations graphiques diverses, ou encore pourront utiliser les nomenclatures "standardisées" comme point de contact entre des fichiers de données, et ce à diverses fins. À titre d'exemple, mentionnons seulement la localisation de clientèle, à partir d'un fichier contenant les adresses postales des clients écrites sans règles particulières, et jumelées, par une programmation spéciale, aux fichiers de nomenclature, et par eux, aux autres fichiers de la base géographique, afin de représenter graphiquement cette clientèle, au moyen d'une table traçante. (re: Figure 3).

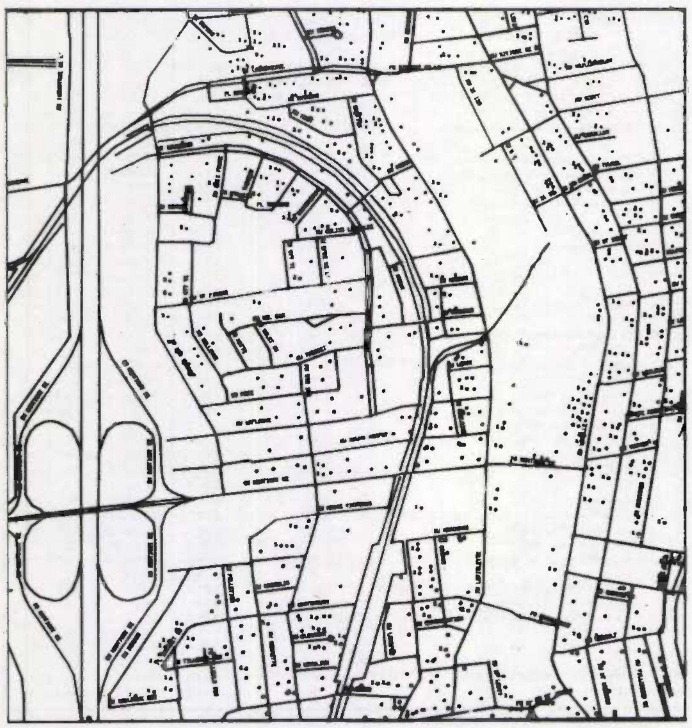

Figure 3 Localisation automatisée de clientèle

Les traitements de ce genre sont d'autant plus facilités par la présence des toponymes populaires ou alias, qui rattachent au nom officiel, par le code de trait, les différentes orthographies ou nomenclatures diverses d'un même trait.

Certaines orthographies, par contre, peuvent être identifiées et "standardisées" par traitement informatique, comme par exemple: St-Paul, dont le St peut être automatiquement traduit en Saint, réduisant ainsi le nombre d'alias (ou toponymes populaires) inscrits au fichier de nomenclature.

L'un des avantages de posséder un fichier de nomenclature construit selon des règles précises, est sans doute la possibilité accrue d'utilisation de ces noms dans des traitements divers. En effet, connaissant les règles syntaxiques qui ont servi à leur inscription, il devient facile de manipuler et d'éditer ces noms, afin de les produire sous une forme approchant plus leur écriture courante, tout en conservant, dans le fichier, une formulation standard mieux adaptée aux traitements informatiques.

Les principaux éléments relatifs à la nomenclature et inclus dans le fichier de nomenclature des traits sont le nom principal, accompagné du préfixe s'il y a lieu, l'orientation et le genre de trait, ces deux dernières informations étant relatives au nom du trait, et non à son genre ou à son orientation physique (sur le terrain). Un code est également ajouté afin de distinguer les noms officiels des toponymes populaires (ou alias).

L'EXPLOITATION DES FICHIERS DE NOMENCLATURE

Divers traitements peuvent s'effectuer à partir des fichiers de nomenclature de la base géographique du système GALILÉE.

Dans les traitements plus traditionnels nous retrouvons évidemment, les listes de ces fichiers, sous différents ordres de tri, soit par ordre alphabétique de nom de trait, soit par ordre de code de trait (re: Tableaux 1 à 4).

Mais grâce à sa forme d'emmagasinement, on peut aussi manipuler les noms et les produire sous des formes dites "éditées", avec le générique (genre) de trait écrit au long (RUE, CHEMIN, BOULEVARD, etc...), l'orientation également écrite au long (EST, NORD-OUEST, etc...), les préfixes replacés en position originelle (dans le fichier: CONCORDE, DE LA; de façon "éditée": DE LA CONCORDE). Les exemples suivants illustrent bien le principe:

Inscription au fichier:

CONCORDE, DE LA	O	BV
VIMONT		PL
0001 IER RANG, DU	NE	CH

Résultat après édition:

BOULEVARD DE LA CONCORDE OUEST ou BOUL. DE LA CONCORDE OUEST
PLACE VIMONT
CHEMIN DU 1IER RANG NORD-EST

Ces formes "éditées" peuvent aussi bien se retrouver dans une liste destinée à des fins de secrétariat ou d'entrée de données, qu'à fournir une source d'information toponymique pour les administrateurs de services à un territoire (re: Tableau 5).

G A L I L E E
B A S E G E O G R A P H I Q U E

NOMENCLATURE DES MUNICIPALITES ET DES LIMITES D INTERET
C.S. DES BOIS-FRANCS (220)

CAT	NOM PRINCIPAL	STAT	LDC	TYP	LIMCOD	MILIEU	MAJDATE
1	BAIE-D'URFE	VT	J	MU	6519	1	751115
1	BEACONSFIELD	C	I	MU	6518	1	751115
1	DOLLARD-DES-ORMEAUX	VT	J	MU	6538	1	751115
1	DORVAL	C	E	MU	6512	1	751115
1	ILE-DORVAL	VT	I	MU	6511	1	751115
1	KIRKLAND	VT	D	MU	6516	1	751115
1	PIERREFONDS	VC	B	MU	6542	1	751115
1	POINTE-CLAIRE	VC	K	MU	6514	1	751115
1	ROXBORO	VT	G	MU	6539	1	751115
1	SAINT-RAPHAEL-DE-L'ILE-BIZA+	P	F	MU	6544	2	751115
1	SAINTE-ANNE-DE-BELLEVUE	VT	I	MU	6521	1	751115
1	SAINTE-GENEVIEVE	VT	I	MU	6541	1	751115
1	SENNEVILLE	V	I	MU	6524	1	751115
2	BAIE D URFE		C	MU	6519	1	751115
2	BEACONSFIELD QUE		I	MU	6518	1	751115
2	DOLLARDDESORMEAU		G	MU	6538	1	751115
2	DORVAL QUE		D	MU	6512	1	751115
2	ILE BIZARD		K	MU	6544	2	751115
2	KIRKLAND QUE		E	MU	6516	1	751115
2	PIERREFONDS QUE		K	MU	6542	1	751115
2	POINTE CLAIRE		J	MU	6514	1	751115
2	ROXBORO QUE		F	MU	6539	1	751115
2	SENNEVILLE QUE		A	MU	6524	1	751115
2	STE-ANNE-DE-BELL		K	MU	6521	1	751115
2	STE-GENEVIEVE		J	MU	6541	1	751115
3	BAIE D URFE QUE		D	MU	6519	1	751115
3	BAIE D'URFE		I	MU	6519	1	751115
3	BAIE D'URFE QUE		F	MU	6519	1	751115
3	BAIE DURFE		C	MU	6519	1	751115
3	BAIE DURFEE		C	MU	6519	1	751115
3	BAIE-D'URFE		H	MU	6519	1	751115
3	BEACONSFIELD 880		K	MU	6518	1	751115
3	BEAUREPAIRE		B	MU	6518	1	751115
3	D D O		K	MU	6538	1	751115
3	D D O QUE		D	MU	6538	1	751115
3	D D ORMEAUX		G	MU	6538	1	751115
3	D D ORMEAUX QUE		A	MU	6538	1	751115
3	D DES ORMEAUX		C	MU	6538	1	751115
3	D DES ORMEAUX QU		G	MU	6538	1	751115
3	D DESORMEAUX QUE		J	MU	6538	1	751115
3	D DO		I	MU	6538	1	751115

Tableau 1 Nomenclature des municipalités par ordre alphabétique

G A L I L E E
B A S E G E O G R A P H I Q U E

NOMENCLATURE DES MUNICIPALITES ET DES LIMITES D INTERET
C.S. DES BOIS-FRANCS (220)

CAT	NOM PRINCIPAL	STAT	LDC	TYP	LIMCOD	MILIEU	MAJDATE
1	HAM-NORD	CT	F	MU	2628	0	751126
1	NOTRE-DAME-DE-LOURDES-DE-HAM	SD	H	MU	2629	0	751126
2	NOTRE-DAME-DE-HA		A	MU	2629	0	751126
3	NOTRE-DAME		D	MU	2629	0	751126
1	INVERNESS	CT	A	MU	2728	0	751126
1	INVERNESS	V	G	MU	2729	0	751126
1	SAINT-PIERRE-BAPTISTE	P	A	MU	2731	0	751126
3	SAINT-PIERRE		H	MU	2731	0	751126
3	ST-PIERRE-BAPTISTE		C	MU	2731	0	751126
3	ST-PIERRE		E	MU	2731	0	751126
1	HALIFAX-NORD	CT	B	MU	2738	0	751126
1	SAINTE-SOPHIE	SD	J	MU	2739	0	751126
3	STE-SOPHIE		E	MU	2739	0	751126
1	SAINTE-JULIE	SD	H	MU	2744	0	751126
3	STE-JULIE		A	MU	2744	0	751126
1	LAURIERVILLE	V	F	MU	2746	0	751126
1	PLESSISVILLE	P	H	MU	2748	0	751126
1	PLESSISVILLE	VT	B	MU	2749	0	751126

Tableau 2 Nomenclature des municipalités par ordre numérique

G A L I L E E
B A S E G E O G R A P H I Q U E

NOMENCLATURE DES TRAITS DU RESEAU ROUTIER
C.S. DES BOIS-FRANCS (220)
PLESSISVILLE P

CAT	NOM PRINCIPAL, PREFIXE	ORIE	GEN	LDC	CODMUN	TRAICOD	MAJDATE
1	ALOUETTES, DES		RU	C	2748	C2C0CC	EOC124
1	ALYSEES, DES		RU	C	2748	C3C0CC	EOC124
1	BEAUDET		RO	J	2748	C4E6EE	EOC124
1	BELLEMARE		RO	E	2748	C53332	EOC124
1	BELLEMARE/BLANCHE		PO	A	2748	C5995E	EOC124
1	BELLEVUE		RG	G	2748	CEE6E4	EOC124
1	BELLEVUE/BOURBON 1		PO	F	2748	C7332C	EOC124
1	BELLEVUE/BOURBON 2		PO	F	2748	C7595E	EOC124
1	BOULEAUX, DES		RU	A	2748	CEE6E2	EOC124
1	BRASSARD		RU	H	2748	C9332E	EOC124
1	CANARDS, DES		RU	E	2748	1140CC	EOC124
1	CHARDONNERETS, DES		RU	J	2748	12E0CC	EOC124
1	COLOMBES, DES		RU	C	2748	1420CC	EOC124
1	CORDON DU 10E RANG		CH	G	2748	15E0CC	EOC124
1	DUBOIS		RO	A	2748	183333	EOC124
1	DUGRE		RU	J	2748	15E6EE	EOC124
1	FORAND		RU	C	2748	22E6EE	EOC124
1	FOUGERES, DES		RU	I	2748	243332	EOC124
1	GARNEAU		RU	F	2748	2550CC	EOC124
1	GERANIUMS, DES		RU	J	2748	2E0CCC	EOC124
1	GLAIEULS, DES		RU	D	2748	265000	EOC124
1	GOUDREAULT		RO	C	2748	270CCC	EOC124
1	GUAY		RO	A	2748	2750CC	EOC124
1	JACINTHES, DES		RU	E	2748	320CCC	EOC124
1	KELLY		RO	A	2748	3420CC	EOC124
1	KELLY/LAURIERVILLE 1		PO	J	2748	3440CC	EOC124
1	KELLY/LAURIERVILLE 2		PO	E	2748	3460CC	EOC124

Tableau 3 Nomenclature des traits du réseau routier par ordre alphabétique

G A L I L E E
B A S E G E O G R A P H I Q U E

NOMENCLATURE DES TRAITS DU RESEAU ROUTIER
C.S. DES BOIS-FRANCS (220)
PLESSISVILLE P

CAT	NOM PRINCIPAL, PREFIXE	ORIE	GEN	LDC	CODMUN	TRAICOD	MAJDATE
1	ALOUETTES, DES		RU	D	2748	C2C0CC	EOC124
1	ALYSEES, DES		RU	C	2748	C3C0CC	EOC124
1	BEAUDET		RO	J	2748	C4E6EE	EOC124
1	BELLEMARE		RO	B	2748	C53332	EOC124
1	BELLEMARE/BLANCHE		PO	A	2748	C5995E	EOC124
3	BELLEMARE		RO	D	2748	C5995E	EOC216
1	BELLEVUE		RG	G	2748	CEE6E4	EOC124
2	ANCIENNE ROUTE 49		RO	F	2748	CEE6E4	EOC216
3	SAINT-GERMAIN		RU	A	2748	CEE6E4	EOC216
3	ST-GERMAIN		RU	A	2748	CEE6E4	EOC216
1	BELLEVUE/BOURBON 1		PO	F	2748	C7332C	EOC124
3	BELLEVUE		RG	I	2748	C7332C	EOC216
1	BELLEVUE/BOURBON 2		PO	F	2748	C7595E	EOC124
3	BELLEVUE		RG	I	2748	C7595E	EOC216
1	BOULEAUX, DES		RU	A	2748	CEE6E2	EOC124
1	BRASSARD		RU	H	2748	C9332E	EOC124
1	CANARDS, DES		RU	B	2748	1140CC	EOC124
1	CHARDONNERETS, DES		RU	J	2748	12E0CC	EOC124
1	COLOMBES, DES		RU	C	2748	1420CC	EOC124
1	CORDON DU 10E RANG		CH	G	2748	15E0CC	EOC124
1	DUBOIS		RO	A	2748	183333	EOC124
1	DUGRE		RU	J	2748	15E6EE	EOC124
1	FORAND		RU	C	2748	22E6EE	EOC124
1	FOUGERES, DES		RU	I	2748	243332	EOC124
1	GARNEAU		RU	F	2748	2550CC	EOC124
1	GERANIUMS, DES		RU	J	2748	2E0CCC	EOC124
1	GLAIEULS, DES		RU	D	2748	265000	EOC124
1	GOUDREAULT		RO	C	2748	270CCC	EOC124
1	GUAY		RO	A	2748	2750CC	EOC124
1	JACINTHES, DES		RU	E	2748	320CCC	EOC124
1	KELLY		RO	A	2748	3420CC	EOC124
1	KELLY/LAURIERVILLE 1		PO	J	2748	3440CC	EOC124
3	KELLY		RO	C	2748	3440CC	EOC216
1	KELLY/LAURIERVILLE 2		PO	E	2748	3460CC	EOC124
3	KELLY		RO	C	2748	3460CC	EOC216
1	KELLY/LAURIERVILLE 3		PO	K	2748	3480CC	EOC124
3	KELLY		RO	C	2748	3480CC	EOC216

Tableau 4 Nomenclature des traits du réseau routier par ordre numérique

G A L I L E E
B A S E G E O G R A P H I Q U E

NOMENCLATURE DES TRAITS DU RESEAU ROUTIER (EDITION REGIONALE)
C.S. DES BOIS-FRANCS (220)

NOMENCLATURE DU TRAIT

GENRE DE TRAIT	PREFIXE	NOM PRINCIPAL, PREFIXE (QUARTIER)
RUE		BUTEAU
RUE		CAMIRE
RUE		CAMPAGNA
RUE		CAMPAGNA
RUE		CAMPUS
AVENUE		CANARDS, DES
RUE		CANARDS, DES
RUE		CANNON
RUE		CANTIN
CHEMIN		CARAVAN VAL BAR, DE
BOULEVARD		CARIGNAN
RUE		CARIGNAN
BOULEVARD		CARIGNAN EST
BOULEVARD		CARIGNAN OUEST
RUE		CARILLON, LE
RUE		CARMEN
RUE		CARON
RUE		CARON
RUE		CARON
RUE		CARRIER
RUE		CARTIER
RUE		CARTIER
RUE		CASTONGUAY
RUE		CAYOUETTE
RUE		CECILE
AVENUE		CEDRES, DES

Tableau 5 Nomenclature éditée du réseau routier (extrait d'un imprimé d'ordinateur)

On retrouve également ces nomenclatures "éditées" dans les représentations graphiques automatisées, où elles présentent la nomenclature rattachée à une représentation de la réalité spatiale (re: Figure 4).

On peut également représenter cette nomenclature sous sa forme d'emmagasinement, c'est-à-dire non "éditée" (re: Figures 5 et 6).

La figure 5 démontre un des problèmes de la représentation graphique automatisée des noms géographiques en présentant une carte d'une partie de la commission scolaire régionale, originalement à une échelle de 1/80 000^{e}. Les problèmes de représentation des noms dépendent de l'échelle d'édition. La programmation, ainsi que certains codes d'emmagasinement en ordinateur, déterminent, selon l'échelle et les options demandées, lesquels des noms de traits seront effectivement tracés.

LES AIRES D'INTERVENTION TOUCHÉES

Actuellement, l'implantation et l'élaboration de la base géographique du système GALILÉE, à l'échelle du Québec, couvre les dix-sept commissions scolaires régionales (C.S.R.) suivantes:

C.S.R. Baldwin-Cartier/Pointe-Claire
C.S.R. des Bois-Francs/Victoriaville
C.S.R. Chauveau/Neufchâtel
C.S.R. Des Monts/Matane
C.S.R. de l'Estrie/Sherbrooke
C.S.R. Honoré-Mercier/Saint-Jean
C.S.R. Lac Saint-Jean/Alma
C.S.R. Lakeshore/Beaconsfield
C.S.R. Lapointe/Jonquière
C.S.R. Le Gardeur/Repentigny
C.S.R. Lignery/La Prairie
C.S.R. Louis-Hémon/Dolbeau
C.S.R. Orléans/Beauport
C.S.R. Provencher Nicolet
C.S.R. Saint-François/Drummondville
C.S.R. Tilly/Sainte-Foy
C.S.R. Vaudreuil-Soulanges/Vaudreuil

En termes de données démographiques, le territoire couvert par ces commissions scolaires touche 27.5% des municipalités du Québec (457/1660), rejoint 23.1% de la population estimée en 1977 (1.41M/6.3M), et s'étend sur une superficie de 100 000 acres, soit 15.5% du Québec méridional.

En termes de données scolaires, ce territoire touche, en plus des dix-sept commissions scolaires régionales, quarante-huit commissions scolaires locales, comprenant environ 830 écoles (bâtisses), pour une clientèle approximative de 200 000 (196 120) élèves transportés.

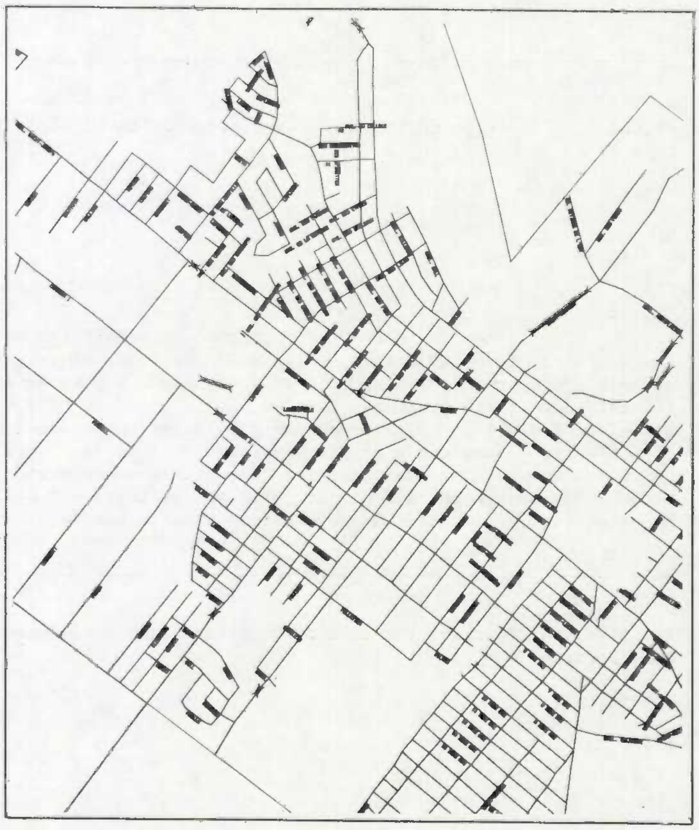

Figure 4 Nomenclature éditée des traits du réseau routier

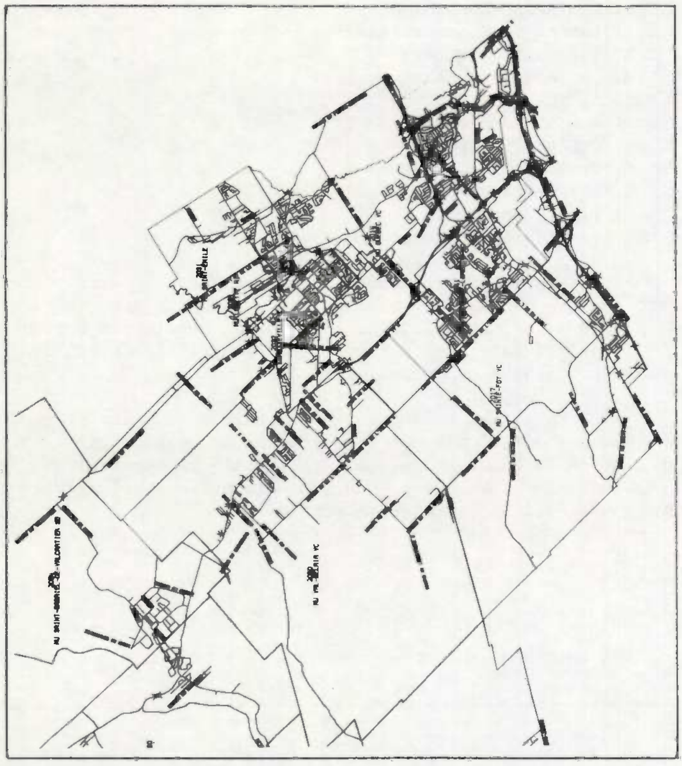

Figure 5 Edition régionale de la configuration du réseau routier

En termes de relevés toponymiques, les différentes bases géographiques du système GALILÉE de ces commissions scolaires comportent présentement environ 30 000 toponymes relatifs aux traits du réseau routier, dont 20 615 de catégorie #1 (officiel) et 9149 toponymes populaires ou parallèles; 15 000 autres devraient s'ajouter au cours de l'année '80-81'. En ce qui a trait aux municipalités, les fichiers comptent environ 1300 noms, dont 457 de catégorie #1 (officiel) et 843 de catégories #2, 3, 4 et 5 (populaires ou parallèles).

Illustrations fournies par le Centre de recherche en aménagement régional.

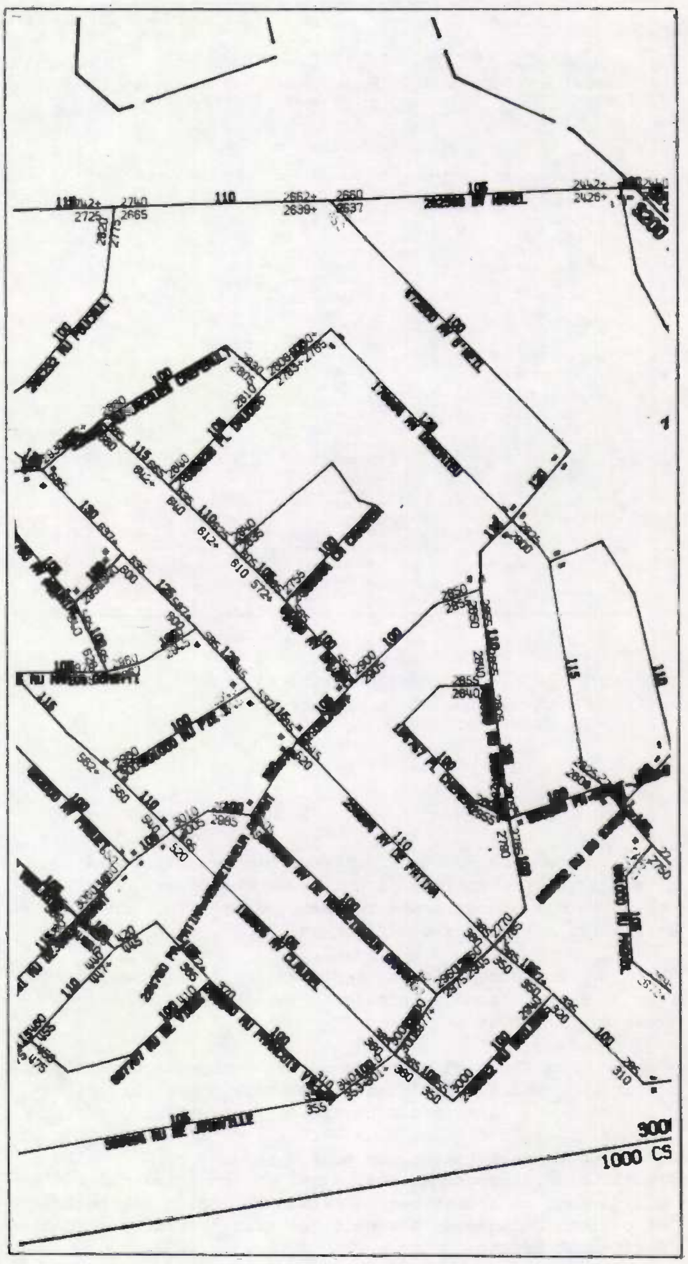

Figure 6 Configuration des traits du réseau routier à l'échelle du 1:5000e avec présence de tranches d'adresses

IDENTIFICATION OF RIVERS FOR STREAMFLOW DATA

D.W. Kirk* and W.J. Ozga**

ABSTRACT. The Water Survey of Canada (WSC) is responsible for the collection, compilation and publication of streamflow, water level and sediment data for approximately 2800 locations across Canada. The official names of rivers and lakes are obtained from the Gazetteer of Canada series and the Guide toponymique du Québec. This paper describes briefly the WSC activities, the uses of water data and the problems associated with the identification of gauging stations.

RÉSUMÉ. La Division des relevés hydrologiques du Canada a pour mission de recueillir, de classer et de publier des données sur le régime des cours d'eau, sur leur niveau et sur les sédiments observés environ à 2800 points à travers le Canada. Les noms officiels des fleuves, des rivières et des lacs sont tirés des Répertoires géographiques du Canada et du Guide toponymique du Québec. Le présent document donne un brève description des activités de cette division, des diverses applications des données recueillies et des problèmes associés à l'identification des stations de jaugeage.

INTRODUCTION

The Water Survey of Canada (WSC) has been collecting and publishing streamflow and water level data on a regular basis since 1908 and sediment data since 1961. The locations where systematic records of these data are obtained, are referred to as "gauging stations".

The present WSC hydrometric network consists of approximately 2450 gauging stations (2060 streamflow and 390 "water level only"); some 2060 stations have been discontinued. In addition, data for 500 active and discontinued stations have been contributed by provincial and private agencies and are included in the WSC data banks. However, some 500 stations operated by the Government of Quebec are not included since it conducts its own hydrometric and sediment surveys and maintains its own data banks. The Water Survey of Canada now operates 104 sediment stations; another 194 stations were discontinued prior to 1979.

Basic data are collected and computed by personnel at seven regional offices in Vancouver, Calgary, Regina, Winnipeg, Guelph, Montréal and Halifax in accordance with national standards established by WSC in Ottawa. The computed data are forwarded annually to Ottawa where they are stored on magnetic tape and made available to users both in publications and in computer-compatible form. In January 1980, all historical daily data to 1978 were converted from imperial to metric units; these past data are now also available on microfiche. In 1979 the total staff engaged in these surveys in the regions and in Ottawa was approximately 350.

Water data are essential for evaluating water supplies for cities, industries, irrigation and recreation; designing structures to control and conserve water; developing guidelines for provincial, national and international administrations regarding licences and permits for water use; determining potential hydropower; studying flood frequency; monitoring water quality; supporting geomorphological studies; and for undertaking other types of water management studies. User needs vary; data may be required for a specific point in time, or for intervals of a few days or months. For design and research purposes the data are most useful after many years of record have been accumulated.

These data are supplied free-of-charge to engineering consultants, universities (libraries, professors and students), provincial agencies (water resources, power, highways and fisheries), federal departments, municipal agencies (waterworks), power companies, railway companies and private individuals. Data publications are provided on a regular basis to some 1132 names on a mailing list which is maintained at WSC in Ottawa. More up-to-date or related data, such as river cross-sectional areas and velocities, are available from the regional offices.

The mandate for collecting data is given under various acts such as the 1909 Boundary Waters Treaty or the Canada Water Act and also in federal-provincial cost sharing agreements.

IDENTIFYING GAUGING STATIONS

Both the regional offices and headquarters make extensive use of gazetteers and topographic maps for identifying bodies of water. The maps are also used for drainage area determination, and the gazetteers are used as the official reference for assigning names.

* D.W. Kirk, Data Review Engineer, Water Survey of Canada, Water Resources Branch, Environment Canada, Ottawa.

** W.J. Ozga, Head, Data Control Section, Water Survey of Canada, Water Resources Branch, Environment Canada, Ottawa.

WSC Station Number

The system introduced over 60 years ago for identifying gauging stations has been retained and is being used also in the storage and retrieval of data. Each gauging station is assigned a unique seven-character identification. The station numbering system begins with the division of Canada into eleven major groups of river basins, each assigned a number. These main divisions are subdivided following heights of land within the division, with a letter assigned to each subdivision. Each subdivision is further divided on drainage basin boundaries and assigned a second letter, e.g. 05BD. Stations in this subdivision are then assigned numbers in chronological order of date of establishment, independent of stream order. Thus, using the seven-character identification system, station 05BD007 is the seventh station established in sub-division 05BD.

WSC Station Name

Each gauging station is also assigned a unique station name which consists of the official name of the watercourse and a location identifier (e.g., a town, river, highway, etc.). These numbers and names are assigned by staff in the regional offices after plotting the station on topographic maps and verifying the names from the appropriate gazetteer. Some examples follow:

02GC010 Big Otter Creek at Tillsonburg
05FE002 Buffalo Creek at Highway No. 44
05BN012 Bow River near the mouth
06DB001 Reindeer Lake at Brochet

Occasionally there are discrepancies between the names shown on the map and in the gazetteer and it is then necessary to contact the Secretariat of the Canadian Permanent Committee on Geographical Names (CPCGN) for confirmation.

Unnamed Watercourse

There are literally thousands of watercourses in Canada which do not have official names. Many of these watercourses are in remote areas of northern Canada, others are the small creeks, ditches, and coulees which make up the millions of tributaries of our surface drainage system. There is often an immediate need for identification of these small streams. An example is the situation which occurred in 1974 in the area surrounding the proposed Pickering Airport, when the WSC was required to supply hydrometric data for drainage study purposes. A network of 10 gauging stations was established on formerly unnamed creeks, and current data were collected and supplied to the design agency. Identification of these sites was required as soon as possible to ensure a minimum of confusion regarding the correct identification for a particular stream.

The same situation frequently occurs when researchers are studying isolated watersheds which require identification of the smallest tributary. Usually the researcher assigns names, e.g., "Inlet No. 1, Lake 239 Outlet, Unnamed Tributary No. 1", rather than wait for an official name to be assigned.

Alphabetical Order

Hydrometric data are published annually in alphabetical order by station name. It is therefore desirable that the watercourses be assigned names in a standard way so that they can be sorted alphabetically by computer and can be found easily. The most common way is where the watercourse is identified by a proper name which consists of a specific term followed by a generic term, e.g., Buffalo Creek, Bow River, Reindeer Lake. The stations can then be sorted in alphabetical order by their proper names. There are exceptions to this standard, e.g., River John, West Branch River Philip, and East River St. Marys. When new approvals are being considered, names such as these should be discouraged.

A name which is unique, but which is identified with reference to another watercourse, should also be discouraged, because it is confusing to users of hydrometric data and more difficult to find in a name search:

e.g. Little Southwest Miramichi River
Northwest Miramichi River
Southwest Miramichi River
Lower North Branch Little Southwest Miramichi River
North Branch South Nation River

It can be confusing to search through a list of "Littles" or "Norths" and it may be ambiguous as to whether "Branch" comes before or after the river name.

For the 30 hydrometric stations operated by WSC in the province of Quebec, the WSC has chosen to identify the names in French but with the generic term in brackets following the specific term, e.g. Richelieu (Rivière), Deux Montagnes (Lac des), and Meach (Ruisseau). In an alphabetical search, this order is more consistent with the English names, where the specific term is usually the first word.

Drainage Area

Ordinarily the WSC computed the drainage area only at the gauging stations, but occasionally there is a requirement to supply the drainage area to the mouth (particularly for rivers flowing into the oceans). The gazetteers provide coordinates for the mouths which are used when computing the drainage areas. However,there is a need for defining how these coordinates are derived, and also to show more precision, e.g. nearest 10 seconds. This is particularly important for large rivers such as the St. Lawrence or the Nelson, each of which gradually opens into a large bay; a change in the location of the mouth could cause a significant change in the drainage area.

Reference Index

In the Surface Water Data Reference Index the WSC publishes a "stream order" index of all gauging stations for which hydrometric data are available. The names of tributaries of the main stem are indented and the indentation indicated by one period; tributaries to these tributaries are further indented using two periods, etc., so that rivers of the same drainage basin can be selected at a glance:

e.g. Fraser River
. Bowron River
. McGregor River
. . Herrick Creek
. . . Framstead Creek
. . . . Muller Creek
. Willow River

In determining the stream order the maps (1: 50 000 or 1:250 000) and the gazetteers (which give the coordinates of the mouth) are generally all that is required, although it is helpful if the gazetteer indicates the name of the watercourse into which the stream flows. However, it is important that each watercourse is identified from its source to its mouth because a change in name usually indicates a tributary and it is necessary to know where that change occurs. An example of this is the St. Lawrence River, which flows from Lake Ontario to the Gulf of St. Lawrence, flowing through Lac Saint-François, Lac Saint-Louis, and Lac Saint-Pierre. Lake Ontario is therefore a tributary to the St. Lawrence whereas the other three lakes are considered part of the river. Also, if the upper reach of a stream above a lake is not identified by name, the lake is often assumed to be a tributary of the watercourse.

The location narrative identifying the name of the watercourse into which a river flows has been discontinued in the Gazetteer of Canada. This information is important to WSC and it is, therefore, strongly recommended that it be restored.

Official Name

It is important that the name for each station reflects the official geographic name, whether or not the station is currently active or discontinued. The supplements to the gazetteers are therefore examined by regional staff to determine any name changes which affect gauging station names. This is a manual task and (considering that there are data for over 5000 gauging stations in Canada) must be performed in each region by someone familiar with the stations of the region.

Occasionally a river has more than one "official" name along its course. This may be a difference in spelling e.g. Kootenai (U.S.A.) - Kootenay (Canada); a change in name e.g. Maynard Coulee (Canada) - Spring Coulee (U.S.A.); the name may change twice e.g. Souris River - Yellow Grass Ditch - Souris River; or there may be both English and French names e.g. Ottawa River - Rivière des Outaouais. These variant forms are confusing to users who are searching for data on a particular river basin, and should be avoided wherever possible.

CONTACT WITH THE CPCGN

It should be noted that as a user of the Gazetteer of Canada series, the WSC has frequent contact with the Secretariat of the Canadian Permanent Committee on Geographical Names to resolve questions concerning the identification of rivers. Many of these contacts are made by telephone and the service rendered by the staff of the CPCGN is very much appreciated. Requests for names for unnamed features are usually submitted by letter, either by the regional offices directly or by headquarters in Ottawa.

CONCLUSION

The Gazetteer of Canada series, its supplements and the Répertoire toponymique du Québec are essential tools used by the Water Survey of Canada for the official identification of gauging stations. Improvements we would like to see in the published volumes and in approval procedures include:

(1) reduction in turnaround time in obtaining new official names for unidentified watercourses;

(2) retention of the tributary identification, e.g., "flows SW into", in the location narrative for watercourses;

(3) adoption of unique names for tributaries, rather than the use of forms like "West Branch";

(4) discouragement of duplication of the same or similar names for different water features; and

(5) more precision in the identification of mouths and headwaters of watercourses.

A NATIONAL GEOGRAPHIC-NAMES DATA BASE

Donald J. Orth*

ABSTRACT. The United States Geological Survey (USGS), in cooperation with the United States Board on Geographic Names (USBGN), has developed a program for the automated processing of geographic name information. The Geographic Names Information System (GNIS) is a geographic names data base which when completed in 1981 will contain about 2 million name entries. Based on the USGS standard map series and other published and unpublished sources, and developed to

* D.J. Orth, Chief, Branch of Geographic Names, National Mapping Division, U.S. Geological Survey and Executive Secretary, Domestic Names Section, U.S. Geographic Board.

meet national needs, GNIS is capable of providing information on a number of data elements including official names, feature class, location of named feature (state, county and geographic coordinates), variant names, map sheet and elevation. Preliminary information is presently available for about 20 states, in the form of Alphabetical Finding Lists, microfiche and magnetic tape.

Under the direction of the Board of Geographic Names the USGS is planning to publish The National Gazetteer of the United States as a series of state gazetteers. An abridged gazetteer of about 40 000 official names covering the major places and features throughout the United States is also to be produced. These gazetteers will be published as chapters of USGS Professional Paper 1200 and changes, additions, and corrections between updated gazetteer revisions will be published in the quarterly reports of the USBGN. The first state gazetteers are to be published in 1981.

RÉSUMÉ. Le United States Geological Survey (USGS) en collaboration avec le United States Board on Geographic Names (USBGN) a élaboré un programme pour le traitement automatisé des renseignements géographiques se rapportant aux toponymes. Le Geographic Names Information System (GNIS) est un système automatisé d'une banque de données toponymiques qui lorsque mis sur pied en 1981 contiendra environ 2 000 000 d'entrées. Il est basé sur l'information tirée des cartes publiées par le United States Geological Survey ou d'autres sources publiées ou inédites est perçu pour satisfaire aux besoins nationaux. Il peut fournir une variété de renseignements: le nom officiel, la classification de l'entité, le lieu (état, comté, coordonnées géographiques), variantes du nom, numéro de carte et élévation. L'information préliminaire, qu'elle soit sur listes alphabétiques, microfiches ou rubans magnétiques, existe pour environ 20 états.

Le United States Geological Survey, sous l'égide du United States Board on Geographic Names se propose de publier The National Gazetteer, un répertoire national, comme un répertoire pour chaque état et un répertoire abrégé comprenant environ 40 000 noms des principaux lieux habités et entités majeures des États-Unis. Ces répertoires seront publiés comme fascicules du USGS Professional Paper 1200 et les changements de noms, additions et corrections entre les mises à jour du Répertoire paraîtront dans les rapports trimestriels du USBGN. Les premiers répertoires des états seront publiés en 1981.

The primary geographic reference system of all people is the use of proper names to identify particular places, features, and areas in the world around them. The use of geographic names illustrates a unique ability to abstract and classify reality, and then symbolize these concepts with sound and visual symbols for purposes of thought and communication with other people. Because of the role they play in controlling the way we think about and personally relate to the world around us, geographic names have strong psychological and cultural significance. They identify areas of cultural and administrative responsibility, define political boundaries, and carry legal weight in determining property, mineral, and water rights. Uniformity in the written use of names and their applications is important, then, for a number of reasons.

The United States Board on Geographic Names was established as early as 1890 to assist in maintaining standard name usage throughout the government of the United States. The success of a national program to maintain uniform usage, however, depends a great deal on effective management of name information. The large amount of interrelated data that must be collected, processed, stored, retrieved, and disseminated requires a system developed around electronic data-processing.

The U.S. Geological Survey, which provides domestic names staff support to the Board, developed such a system and is in the process of building a computerized national geographic names data base. Referred to as the Geographic Names Information System (GNIS), the system (Figure 1) is capable of providing basic information for about two million names and their related features in the United States, when the file is completed early in 1981. Developed in cooperation with the Board on Geographic Names to meet national needs, the data base will:

(1) assist in establishing and maintaining official names usage throughout the federal government in cooperation with state and local government and the public;

(2) provide an index of names found on federal, state, and private maps;

(3) eliminate duplication and the need to spend large amounts of money and time by government agencies, industry, and institutions to organize similar basic data files for specific needs;

(4) provide for standardization of data elements and their coded representation for use in information interchange within the information processing community;

(5) meet federal public information requirements.

The system is designed to be used by and provide name information not only to federal agencies, but also to state and local governments, industry, commerce, scientific and educational institutions, and the general public. The data base is managed by the Branch of Geographic Names in the newly reorganized National Mapping Division in the Geological Survey. In general, the system serves two kinds of users:

(1) those who will use the information for reference purposes;

(2) those who will use the file as a geographic base or subset for larger specialized data files.

The first group is interested in learning official names or locating places and features and can use the planned National Gazetteer and the Alphabetical Finding Lists (Figure 2) that are furnished mainly on a state basis. Special list-

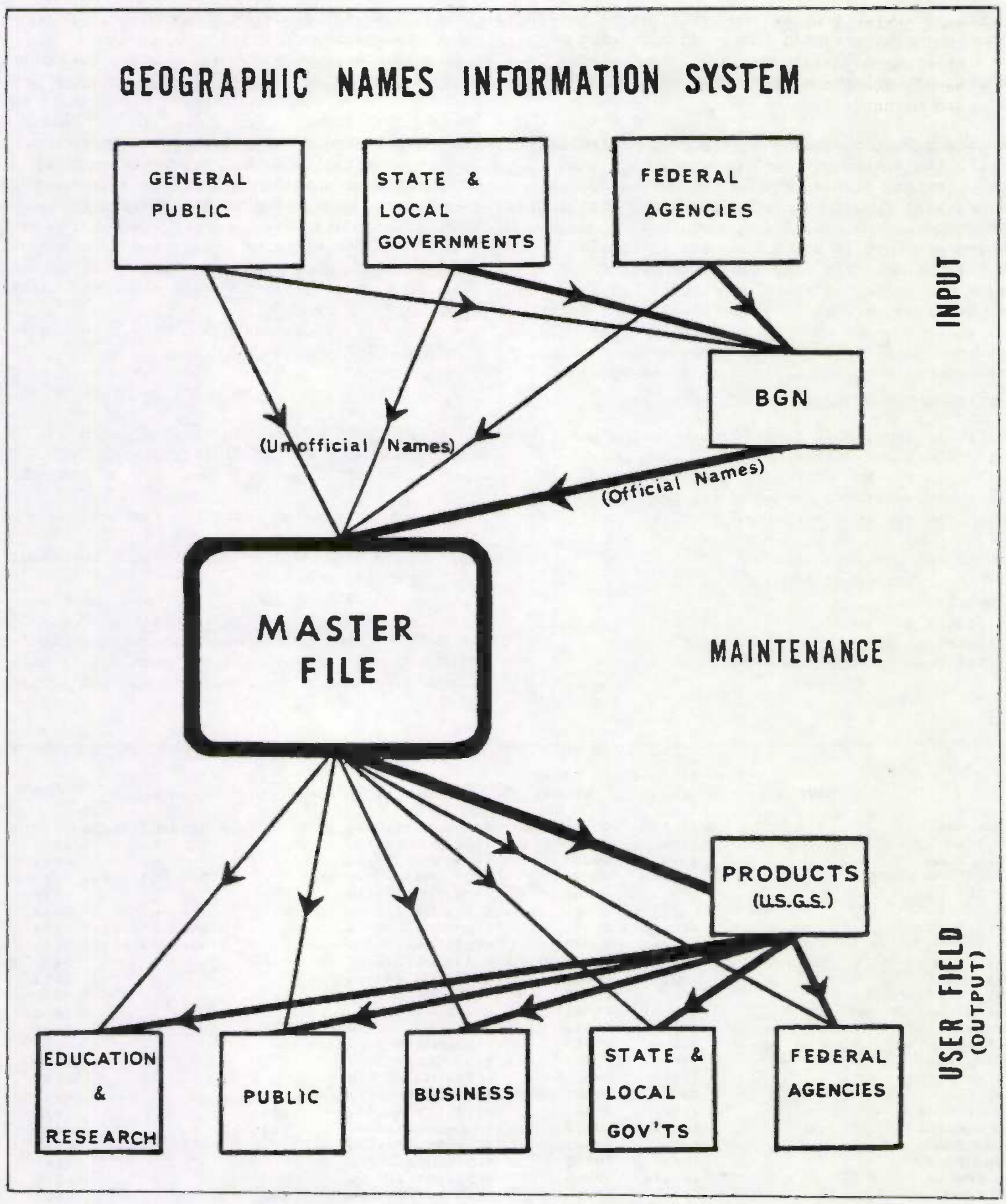

Figure 1 The Geographic Names Information System: diagram showing the flow of data and information into and from the system.

ings sorted by area, feature, or elevation also can be furnished. For example, an organization may need a listing of named water features in Appalachia, or a list of populated places in the United States with an elevation less than 7Q0 feet.

The second group of users needs special categories of names for particular areas, but for purposes other than direct reference. Several states plan to use GNIS as a geographic base for natural-resource inventories and control. Manufacturing, transportation, and tele-communication companies have expressed interest in using the data in special-purpose files requiring accurate name and location information. These files may relate to shipping matters, rate structures, user profiles, or facility locations. Important to cartographers is the use of GNIS in support of a digital cartographic data base with applications in computer mapping, particularly for name selection and placement, and type and symbol generation.

The national geographic-names data base includes, or will include, primary information for all known places, features, and areas in the United States identified by a proper name. These include natural and cultural features. The kinds of cultural features represented in the file are:

(1) populated and non-populated places;
(2) major and minor civil divisions;
(3) dams and reservoirs;
(4) national, state, and local parks;
(5) military installations;
(6) airports, railroad crossings and sidings;
(7) major shipping facilities;
(8) reservations;
(9) cemeteries, churches, schools, buildings, and hospitals.

When completed the data base will contain information on official standard names used in the United States, i.e., names established as official by the Board on Geographic Names, and those established by legislative and administrative action at the federal and state levels. Because of this, the highest priority is given to data base integrity (quality of input and update) and data base security (prevention of unauthorized changes). Besides official names, the data base will include several categories of minor entities, such as cemeteries and churches, for which official names are often difficult to determine.

It is intended that when practical every record include the following:

(1) official name, its spelling, and written form;
(2) kind of feature;
(3) state and county in which located;
(4) geographical coordinates, which locate features and the extent of linear features such as streams, canyons, and valleys;
(5) map or maps on which the feature can be found;
(6) elevation;
(7) date when name and application were made official; and
(8) variant names.

Many records in the system already include other kinds of data related to the names and associated features. Information such as name origins and meanings is important to many people and will be added whenever time and availability permit.

NAME	FEATURE CLASS	STATE COUNTY	COORDINATE	BGN	ELEV FT	SOURCE	MAP
Keach Brook	stream	09015 44007	415500N0714047W	1972		415713N0714623W	0029
Keach Pond	lake	09015	415813N0720516W				0027
Keating Pond	lake	09001	412303N0731520W				0077
Keating Pond Brook	stream	09001	412410N0731527W	1965		412305N0731520W	0078 0077
Keefe School	school	09009	412023N0725611W				0095
Keeler Cove	bay	09005	414229N0731357W				0048
Keelers Brook	stream	09001	410458N0732650W			410556N0732819W	0114
Keene Brook	stream	09013	415416N0721645W			415635N0721712W	0025
Keeney Cove	bay	09007 09003	414310N0723746W	1941			0052 0053
Keeney Point	cape	09007	414334N0723049W				0052
Keeney Street Pond	lake	09003	414522N0723236W				0038
Keeney Street School	school	09003	414523N0723220W				0038
Keeny Cove	bay	09011	412102N0721048W	1942			0101
Keep Swamp	swamp	09005	415949N0731902W				0017
Keighley Pond	lake	09007	413322N0723301W				0068
Keley Pond	lake	09009	412140N0724201W				0097
Kelley Pond	tank	09005 09009	413055N0730921W				0063
Kellners Pond	lake	09001	412451N0732644W		447		0076
Kellogg Corners	locale	09005	415042N0731050W		1016		0032
Kellogg Pond	lake	09001	410750N0732553W				0107
Kellogg Pond	lake	09005	415716N0732101W				0017
Kelly Corner	locale	09005	415242N0731225W				0018
Kelly Pond	tank	09005	414557N0730451W		867		0034
Kelly Pond Brook	stream	09005	414505N0730349W			414559N0730451W	0034
Kelly Ponds	lake	09015	415234N0715930W		483		0020
Kelly Slide	cliff	09005	413456N0732704W				0061
Kelsey Brook	stream	09005 36027	415706N0733000W			415743N0732926W	0016
Kelsey Hill	summit	09007	412212N0722648W				0099
Kelsey Island	island	09009	411500N0725113W				0096
Kelsey Point	cape	09007	411525N0723027W	1938			0098
Kelseys Pond	lake	09007	413609N0723409W				0068

Figure 2 Alphabetical Finding List: part of sample page from the Connecticut Geographic-Names Alphabetical Finding List. Each state list includes both a numerical and alphabetical index to the topographic maps covering the state and an index to the Federal Information Processing Standards (FIPS) state-county codes.

The storage, maintenance, and retrieval of information in the data base are presently handled by an information processing program called GIPSY (General Information Processing System), developed by the University of Oklahoma, and presently run on IBM System 370/155 computers. GIPSY can assemble and process large data collections composed of numerical, codified, or natural-language information. Variable-length records are accommodated and data elements can be readily updated or new information added. More data elements can and probably will be added in the future. The system does not establish preconceived relationships between data collected and stored, and allows users with no programming experience to easily access the file through catalogued routines. Any data element or set of elements can be searched and retrieved.

From GNIS we can furnish alphabetical and topical listings in the form of computer printouts and magnetic tapes (9-track 1600-BPI EBCDIC codes) for states and areas within states. Listings in microfiche are also available.

Information from the system also will be available in a way that will fulfill a longstanding effort at the Geological Survey. As part of its mapping program, the Geological Survey began cataloging names as early as 1892. It was planned to produce a series of state gazetteers "designed as an aid in finding any geographic feature upon the atlas sheets published by the Geological Survey". Compilation was initially done by Henry Gannet, chief topographer, under the direction of John Wesley Powell, second Director of the Survey. Gazetteers for twelve states, Puerto Rico, Cuba, Territory of Alaska, and Indian Territory (Oklahoma) were published between 1894 and 1906. Topographic mapping, however, is a slow process, and inadequate map coverage restricted the gazetteer program.

The Geological Survey, by authority of the U.S. Board on Geographic Names, now plans to publish the National Gazetteer of the United States on a state-by-state basis. All volumes of the National Gazetteer will be published as chapters of USGS Professional Paper 1200, and each will be identified by state name and date. Changes, additions, and corrections to entries in each state volume will be available through the official reports of the Board on Geographic Names.

The Geological Survey began preliminary work on the data base in 1976 at a time when published large-scale topographic maps were available for more than 70 percent of the country. In a pilot project, the University of Oklahoma collected and put into machine-readable form the names published on the Survey's maps covering the states of Kansas and Colorado. In the fall of 1978, Automated Datatron, Incorporated (ADI) of Washington, D.C. was awarded a contract to complete the rest of the United States. The coding of the first states on the new contract was delivered early in 1979. While initial work was slow in order to develop procedures and accuracy, ADI is now completing about four states per month, with completion of the contract scheduled for October 1, 1980. Work is being monitored for completeness and accuracy by the Branch of Geographic Names (Figures 3 and 4). The main steps in building the data base are as follows:

(1) *Map Acquisition and Numbering.* Acquisition of complete topographic map coverage published by the Geological Survey on a state-by-state basis. Each map of a state set is given a unique sequential number beginning with "1" in the northwest corner of the state and numbering latitudinally from west to east.

(2) *Map Annotation.* After delivery of the maps to the contractor, a geographer identifies each named place and feature, gives each a unique number within the topographic map and annotates the extent of certain features such as streams. Each feature is related to all maps on which it appears.

(3) *Data Entry.* A keyboard operator registers the four corners of the map and digitizes the location of each feature in order by feature number and other required data

Photo: Donald Orth

Contractor personnel annotating map features for key to disk operation.

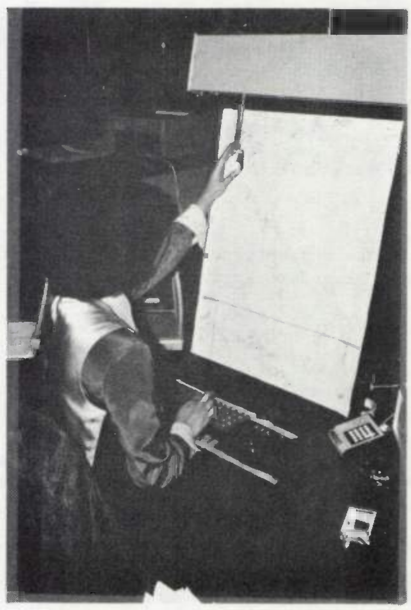

Photo: Donald Orth

Contractor placing digitizing cursor in key to disk operation to input name and geographic coordinates in degrees, minutes and seconds.

Figure 3 Work of contractor personnel in building the Geographic Names Information System (GNIS).

Photo: Donald Orth

USGS personnel mounting tape to provide instructions to generate the computer overlay map used in verifying contract data.

Photo: Donald Orth

USGS personnel checking contract data. Names and positions computer plotted on a clear overlay can be visually compared with the topographic map; other attributes are checked on the computer printout.

Photo: Donald Orth

Some corrections are keypunched to up-grade the data base; other updates may be made by using magnetic tape.

Figure 4 USGS personnel monitoring contract work completed for the Geographic Names Information System (GNIS).

onto magnetic tape; the X-Y coordinates are converted directly into geographic coordinates.

(4) *Process and Editing.* A series of programs checks accuracy of data and sorts information into a required record format.

(5) *GIPSY Conversion.* The contractor's tapes are delivered to the Geological Survey and the data are put into the GIPSY storage and retrieval system.

(6) *Data Monitoring.* Names data for specific 1:24 000-scale maps are pulled out of the system in printout form. Also, names and positions are plotted on a mylar overlay for immediate visual comparison with the source maps.

(7) *Editing Update.* Data are checked against the Board on Geographic Names files and updated to include changes, additions, and corrections that may have occurred since the topographic maps were published. The Board's decision dates, as well as any variant names, are added.

(8) *Final Formatting.* Information is formatted into a one-line 132-character entry that displays information from each data element (except variant). Space requirements necessitate limiting the extent of the information printed in some data elements. A full printing of all information in every data element may be obtained by a special search at a somewhat higher cost. In addition, special searches are available and are limited only by the data in the system and the imagination of the user.

Several reference systems have been developed which identify points and areas on the earth's surface. The system based on geographic names, however, differs from the rest in that it is tailored to be used by all people. It includes not only locational, but also place and feature concepts which can be communicated in everyday language. GNIS will give us the opportunity to extend the usefulness of geographic names by providing a standard information base for a multitude of activities.

AUTOMATION OF FOREIGN PLACE NAMES AT THE UNITED STATES DEFENSE MAPPING AGENCY

William E. Opalski*

ABSTRACT. The United States Defense Mapping Agency (DMA) has determined from cost analysis that considerable savings can be achieved by automating the Foreign Place Names process. The United States Army Engineer Topographic Laboratories (ETL) has been tasked to perform the research and development associated with the project. The prototype Semiautomated Type Composition Console System, the Automatic Type Placement System and the Electron Beam Recorder are briefly described. The technical issues related to a requirements determination and a file structure for a Foreign Place Names Information System are discussed. The proposed solution to solving the problems of selecting a data base management system for names and an input CRT terminal capable of keying diacritical marks as well as alphanumerics are stated.

RÉSUMÉ. Le United States Defense Mapping Agency a déterminé, par suite d'une analyse des coûts, qu'il est possible de réaliser des économies considérables en informatisant les toponymes étrangers. Les United States Army Engineer Topographic Laboratories se sont vues confier la tâche d'exécuter les travaux de recherche et de développement associés à ce projet. On y décrit l'appareil de typographie semiautomatique à console et l'appareil de typographie automatique ainsi que l'enregistreur à faisceau d'électrons. On étudie les questions techniques associées à la détermination des besoins et à la composition des dossiers, pour un système d'information sur les toponymes étrangers. On expose la solution au problème du choix d'un système gestion de la base de données toponymiques et la nature d'un terminal TRC capable de tenir compte de signes diacritiques et alphanumériques.

INTRODUCTION

The main concern of this meeting, automation of geographic names, is one shared by a number of institutions and agencies in the United States. The topic has been on the agenda of many meetings of commercial map-making houses, colleges and universities, state and local agencies, and federal offices and bureaus for perhaps twenty years. All of these organizations have been seeking ways of modernizing cartographic and related operations by exploiting the advantages of the computer age. Over the past few years, there has been a groundswell of interest in the automation of geographic names at all levels, including the international concern expressed by the International Cartographic Association and the United Nations.

The Defense Mapping Agency (DMA), which is the agency of the United States responsible for producing maps, charts, and geodetic products for the armed services and for various civil purposes, has investigated names automation on several occasions in the past. This interest was based on the premise that the advances made in other aspects of map and chart production by automated processes should be matched by equal advances in the automation of names. At the same time, the United States Board on Geographic Names (BGN), which has close ties with DMA, saw that its responsibilities for providing standardized names information for official U.S. purposes would be better fulfilled if automation could be applied. Even though some progress had been made through the production of BGN gazetteers from magnetic tapes, the focus had been on automated production of names lists with little concern for creating a names file for other purposes.

The DMA-BGN relationship stems from the fact that DMA provides staff support to BGN foreign-names programs. At the DMA Hydrographic/Topographic Center near Washington, D.C. a unit of some 20 names experts in the Toponymic Branch carries out various BGN programs. They maintain a card file of some 4.5 million names (in the Department of Defense Foreign Place Names Library), undertake research in preparation of BGN gazetteers (of which about 170 have been published, a number of which need revision), and respond to some 900 inquiries each month for names information. This work also benefits DMA mapping and charting programs in various ways. Another DMA unit (the Names Application Branch) is engaged in straight names application for map revision purposes, without developing data for file purposes.

Recently, DMA renewed its attention to the matter of names automation for several reasons. First, the experience of past efforts pointed toward solutions of problems that had stymied earlier plans. Second, technology has improved to the extent that requirements for treatment of such items as, for example, diacritics could be met. In addition, the costs of con-

---

* William E. Opalski, Project Engineer, Mapping Developments Division, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia.

tinuing with strictly manual methods of names processing were becoming prohibitive when balanced against the growing needs for current names information in a world where names were proving to be as unstable as ever. These factors demonstrated that a specific course of action was required and that an in-house capability in names automation was not only desirable but also necessary. Accordingly, DMA has undertaken a program designed to create a fully automated system of names treatment.

THE REQUIREMENTS

The Defense Mapping Agency has determined through cost analysis that considerable savings can be achieved by automating the geographic names process. It has also determined that other benefits such as reducing manpower requirements, establishing compatibility with existing automated cartographic systems, and increasing the names file utility can be derived from automation. DMA has therefore decided to build a Foreign Place Names Information System (FPNIS) to meet the following requirements:

(1) have an interactive edit and query capability;

(2) update in batch or on-line;

(3) retrieve in batch or on-line;

(4) be capable of performing ad hoc queries by non-programmers;

(5) be capable of entering and storing alphanumerics and appropriate diacritical marks;

(6) generate, through established query formats, output files that can be used in application programs for type placement both for mapping and gazetteers;

(7) be capable of answering interactive queries from users outside of DMA;

(8) provide data base management software functions such as security, report generation, data independence, and recovery;

(9) be flexible to changing requirements;

(10) be expandable up to 26 million names;

(11) be able to key or query by:

- political entity (country);
- alphabetical order (listing);
- individual entry;
- approved names or variant names;
- geographic or grid coordinates;
- designation (mountain, river, city, bay);
- map sheet/chart number; and
- population.

(12) be capable of handling a record of up to 2000 characters.

A dedicated minicomputer system seems to be the solution to these requirements.

DMA has tasked the U.S. Army Engineer Topographic Laboratories (ETL), a research and development (R&D) laboratory, which supports DMA in developing a variety of automated cartographic systems, to do the R&D tasks associated with building the FPNIS. ETL is focusing on four major areas:

(1) input hardware/software systems for the names file (which requires diacritical marks in several foreign languages);

(2) file structure;

(3) data base management system software; and

(4) names placement functions.

The work effort will be in the form of studies, demonstrations, pilot programs and prototype systems. DMA will incorporate the technology and information acquired from R&D into an automated toponymic data base system (FPNIS) that can provide users with up-to-date foreign place names and information about the names. This approach will establish the practicality and feasibility of a multitude of special requirements before they are actually implemented in a production data base system.

INPUT HARDWARE/SOFTWARE

One of the first problems encountered in the automation of names was to find a CRT terminal and keyboard capable of keying diacritical marks associated with romanized geographic names in various foreign languages. A study of DMA requirements has indicated that there are 343 roman script-diacritical mark combinations used in map and gazetteer production.

After a survey of terminals on the market, ETL decided that a programmable intelligent terminal called the "Smart ASCII" manufactured by the ECD Corporation of Cambridge, Massachusetts has the essential features needed for developing fonts of diacritical marks, which can be displayed on a CRT screen. The system is designed to allow the user to define his own characters for foreign languages. In addition, the programmable features provide maximum flexibility for developing screen formats and for changing from one screen format to another. New characters (diacritics) can be added at any time by filling a matrix on the screen, while relegendable key caps[1] make it simple to incorporate the new diacritics into the keyboard. The diacritical marks appear on the video screen in the proper place in relation to the alpha characters. The terminal also has a text formatting and an interactive edit capability with familiar features such as insert, modify, delete, swap, move, and search. Editing programs can be written by the user.

The terminal can be attached to a minicomputer such as a PDP 11/70 and used to load and access a data base management system or it can be used as a stand-alone unit. ETL intends to use the terminal both ways.

ETL has procured two terminals and is currently developing an input format for the Foreign Place Names File that includes a series of menu prompts[2] which allows the operator to fill in the "blanks" of a data record on the screen. The input software will allow the record format to be converted to a gazetteer page format on command.

The ECD terminal has, as a peripheral aspect, a serial matrix printer manufactured by the Florida Data Corpora-

---

1 Key tops which can be physically interchanged to show different alpha or numeric characters (e.g. "c" might be replaced by "č").

2 A list of statements or messages, usually in the form of questions, displayed on the CRT screen to provide the operator with the function options of the system.

tion of Melbourne, Florida. The printer is engineered to share the memory of the terminal and therefore will print out exactly what is shown on the terminal video screen. The printer is used to print proof copy, but has 15 x 16 matrix which can be employed as an option to produce fully formed characters and could be used to produce an interim copy of a gazetteer.

The names input on the terminal when used as a stand-alone unit are stored on a Biflex dual floppy disk drive system. The floppy disk has an IBM 3740 format which ensures interchangeability with other systems; and in particular, DMA's MULTISET III which is an intelligent typesetting input minicomputer system built by the Alphatype Corporation. The MULTISET III system specifications includes a Computer Automation LSI 4/30 16 bit word minicomputer with 96K bytes of memory, a CDC 80 megabyte fixed head disk, a Shugart 800 Series floppy disk, intelligent terminals with programmable character set capability, and a nine track tape drive. The MULTISET III is on-line to a Photon Mark IV Pacesetter typesetting system at the DMA Hydrographic/Topographic Center.

In the initial development, the DMA will create a file of foreign place names on an ECD terminal as a stand-alone unit. The floppy disks with names will be transported to the MULTISET III and the file transferred to the hard disk. A copy of the file will be transferred by magnetic tape to the data base management systems at ETL for further experiments. DMA Hydrographic/Topographic Center will extract names from the file on the MULTISET III and print them on the Photon in a gazetteer format. Archival storage will be magnetic tape.

FILE STRUCTURE

The proposed content of the digital Foreign Place Names File will have the following data elements:

(1) U.S. Board on Geographic Names (BGN) approved name;

(2) BGN approval date;

(3) BGN approved conventional name;

(4) Department of Defense (DOD) approved name (where different from BGN name);

(5) variant name;

(6) substandard name approved by BGN;

(7) Joint Operational Graphic (JOG) ID number or map number;

(8) body (island, continent);

(9) country;

(10) administrative division;

(11) geographic coordinates to nearest second;

(12) Universal Transverse Mercator (UTM) grid coordinates;

(13) designator;

(14) measurement of the feature;

(15) population;

(16) hierarchical code;

(17) information source; and

(18) remarks.

It is anticipated that data elements will be added or deleted as data base experiments progress.

DATA BASE MANAGEMENT SYSTEM

ETL has a keen interest in commercially available Data Base Management Systems (DBMS) that can be used to manage cartographic data, including names. It is felt that a well documented, debugged, production-proved DBMS will be more economical to implement in a production environment than a custom-made system. But since it is not well understood how the commercial systems will work with cartographic data, ETL recently procured three DBMS systems which represent the hierarchical, relational, and network philosophies and have installed them on a PDP 11/70 minicomputer for evaluation. The intention is to test the DBMS with various cartographic data bases in a series of experiments. These systems are the Data Retrieval System (hierarchical), Oracale (relational), and DBMS-11 (network/codasyl). Another DBMS system, developed for the U.S. Air Force, called the Search and Retrieval Processor, which has a geographic search processor, will be installed on a PDP 11/45.

The names file created on the DMA intelligent terminal will be tested in each of these systems to determine which DBMS structure is optimal. A second ECD terminal will be procured to load and access the file on the DBMS. In addition, DMA's extraterrestrial file (analog) is currently being loaded into Oracale. The plan is to load seven thousand records into the system as a pilot project. A file of extraterrestrial names was chosen as an interim study because it has few diacritical marks and can be accessed with a regular Tektronix terminal. The relational data bases are designed to provide the capability for unstructured queries. There is concern whether the relational system will work efficiently with a file of several thousand names since this DBMS system overhead is very high.

The largest part of the DBMS text effort will be exercised with the Search and Retrieval Processor (SARP). ETL will use the second ECD terminal primarily on the PDP 11/45 to operate the SARP. A capability to load SARP from the terminal will be developed along with a software input processor to load the DOD Foreign Place Names File from DMA. ETL also plans to integrate a digitizing table into the PDP 11/45 system so that name coordinates can be digitized from existing charts and maps. One of the main objectives will be to test the geographic search processors on the SARP as these search routines are a major requirement which cannot be satisfied with many commercial systems.

NAMES PLACEMENT FUNCTION

ETL has developed for the DMA Aerospace Center a semiautomated Type Composition Console system that is used to locate (manually) and digitize the rectangular areas on a map where names are to be placed. The system has two 50 x 60 inch plotter/digitizer tables on which the operator places a map work sheet. The names selected for placement are listed on a CRT screen that sits next to the table. The operator can select a place for a listed name with a cursor, and the plotter mechanism will draw a rectangle which will precisely fit the size of the name (length and weight) in place on the map work sheet. The rectangle coordinates along with other information such as point and font will, as a separate process, be read into an Automatic Type Placement System which is independent of the console.

Additional equipment in the Type Composition Console system are a Computervision Graphics Processor minicomputer with 104K of memory, a FORTRAN compiler, 2 magnetic 9 track tape units (800/1600 BPI), an 80 megabyte disk, 2 input terminals (one ECD intelligent terminal) for listing names, and a card readers. The console has interactive edit capabilities which can be employed at the CRT input terminals or at the CRT plotter table terminals. The system can perform three operations simultaneously, i.e., operate both tables and one input terminal, but can be expanded to support twelve operations simultaneously.

ETL has also developed an Automatic Type Placement system which has been installed at both DMA Centers. In this system, a cathode-ray tube printhead has been retrofitted to a precision flatbed plotter in place of the pen holders. Alphanumeric characters and symbols at 4-72 point size are drawn on the CRT face which exposes photographic film placed on the plotter table. Character and symbol data are stored in digital form on a magnetic disk in two master sizes and in a variety of cartographic type styles. Entire names/symbols or groups of names/symbols may be exposed, at or near full scale during each 2-inch step of the plotter table.

Equipment to drive the CRT consists of a Tektronix 4014 graphics terminal, magnetic disk units, a magnetic tape unit, and a PDP 11/45 minicomputer. A type-placement capability has also been included in an Electron Beam Recorder (EBR) development under contract with Image Graphics, Inc., which will output all map names and symbols at an appropriate size reduction using up to a 5" x 8" film format. Input to the CRT and EBR systems is a magnetic tape of names/symbols data consisting of names/symbols, type fonts, location coordinates, and orientations. A computer receives the input names/symbols and date, calls out the proper symbol/type styles from the digital type style and symbol library on the disk file, and drives the CRT or the EBR printer to plot onto film.

The names file generated from the type placement file will not be stored in the Foreign Place Names Information System at DMA Hydrographic/Topographic Center.

CONCLUSION

This paper outlines the plans being implemented by DMA to introduce automation into its names program. It discusses some of the problems identified to date and it deals with technologies being considered as solutions to the problems. It also stresses the need to proceed in phases, each one of which will be tested prior to advancing to the next. In describing DMA goals for automation, the paper above all demonstrates the commitment to develop a capability to manage names information in a fully automated environment. This capability will, in turn, enhance the mission of BGN to provide standardized names information to users at all levels.

Photo: Mike Kelly, Information EMR

SYMPOSIUM SPEAKERS/ORATEURS DU COLLOQUE:

(Left to right/de gauche à droite): William Opalski, Stewart Kingsbury

AUTOMATION OF GEOGRAPHICAL NAMES IN MICHIGAN

Stewart A. Kingsbury*

ABSTRACT. A computerized data model has been developed for the automation of geographical names in Michigan. The model utilizes a geographical locator number for each name (e.g., Marquette County: 125.1 126.26 127.103) and then computerizes historical, geographical, folkloric and linguistic data around each of the more than 5555 geographical names of the state's Upper Peninsula. Included are more than 70 types of place name designations, such as county, township, city, village, mountain, lake, river, bay, point, forest, park and mine.

RÉSUMÉ. Un modèle informatique vient d'être mis au point pour le traitement des noms géographiques du Michigan. Le modèle désigne chaque nom par un numéro localisateur géographique (par exemple le comté de Marquette vient 125.1 126.26 127.103), puis donne les données historiques, géographiques, folkloriques et linguistiques qui concernent les quelque 5555 noms géographiques de la haute péninsule de l'état. Sont compris plus de 70 types de désignations de lieux comme comté, canton, ville, village, montagne, lac, rivière, baie, pointe, forêt, parc et mine.

INTRODUCTION

In 1976 a computerized data center for the *Place Name Survey of the United States* (PNSUS) was established under the direction of the writer, at Northern Michigan University in Marquette. Conducted under the joint auspices of the American Name Society and the United States Board on Geographic Names, the first efforts of PNSUS were concentrated on automating the names of the Upper Peninsula of Michigan.

DATA ACQUISITION

A team of research students extracted 5555 names from the United States Geological Survey 7.5 and 15 minute quadrangle maps and entered each of them on a *yellow locator card* designed by Byrd Granger[1] (see Figure 1). This information was supplemented with historical, linguistic and folkloric data. Other references were consulted so that there is now a listing of approximately 11 800 names relating to places, railway points and land and water features in and adjacent to the Upper Peninsula.

Based on Donald Orth's computerized data model[2], *special research codes* were chosen from 000.00 to 999.00. The numbers 005.00/049.00 were assigned to oral history subject categories; 050.00/099.00 to linguistic geography: 125.00/499.00 to name classification: man-made and God-made; and 500.00/999.00 to folkloric classification.

Data Increments for Collation of Names Information

(1) Name: Complete place name is recorded according to standard alphabetizing procedure.

(2) Genre: The appellative or generic element, including prefix or suffix is listed separately for research purposes.

(3) Spec: Special research codes in numeric form are listed and separated by a semicolon. Present categories: 000/004--computer instruction; 005/049--special group listings; 050/099--language information; 125/499--name classification; 500/999--unassigned; 1000/2100--date of name origin; 5-digit number records local relief information.

(4) BGN: The date of a Board on Geographic Names decision for the name, if any.

(5) Desig: A standard term identifying the kind of feature the name applies to.

(6) Loc.: The state and county in which the feature is located are identified according to the Federal Information Processing Standards (FIPS) numeric code published by the National Bureau of Standards.

(7) Map: The National Topographic Series map or maps on which the feature is located is designated by identifying numbers.

(8) Coord: Geographical coordinates to define the precise location of the feature are given as specified in the foreword of Board of Geographic Names decision lists.

(9) Pop: If applicable, the population of a place or area is recorded with date and source of information.

---

* Dr. Stewart A. Kingsbury, Professor, Department of English, Northern Michigan University.

1 Granger, Byrd Howell, *Handbook of Establishing a State Center for a Survey of Place Names*, Department of English, University of Arizona, (no date), p. 6.

2 Orth, Donald, *Computers and place names: a new challenge*. USGS, Reston, 1970, p. 9-10.

Place Name — Map source

Latitude — Longitude — Township — Range — Quarter

Elevation

Category — (river, mountain, plain, city, etc.) (dry river bed, BE AS SPECIFIC AS POSSIBLE)

Remarks:

Researcher — "B"

Figure 1 Byrd Granger Yellow Locator Card

(10) Size: Linear and areal dimensions of the feature are recorded if possible.

(11) Elev.: Elevation of the place or feature is given in feet above sea level.

(12) Drain: An alphanumeric code established by the Office of Water Data Coordination, U.S. Geological Survey, indicates the drainage basin in which the feature is located.

(13) TR: U.S. land survey information by section, township, range, and principal meridian is recorded when applicable.

(14) Descr.: Natural language description of the feature records general information not found under any other label.

(15) Var.: All known variant names applied to feature are listed alphabetically for cross referencing.

(16) Hist.: The history of the name is recorded, if known. Who gave the name or first used it. When, and why the particular words contained in the name were chosen. Folk etymologies are also recorded and so identified. All information listed is referenced to a source. If the place name is also the source of a geologic name, the formal geologic name, author of the name, and date of first publication are given.

INPUT: PREPARATION AND KEYPUNCHING THE DATA

The computerized program fixes all historical, linguistic and folkloric data to fixed points described by latitude and longitude and by township, range and section. The first information entered is the *cross reference number* derived from each *yellow locator card*. This detail is arranged as follows:

(1) place name;
(2) map source;
(3) latitude and longitude;
(4) township, range and section;
(5) elevation;
(6) name classification category;
(7) population from the latest census figures.

The next input entry is historical, linguistic and folkloric data, designated by a *précis number* describing the location fixed by the *cross reference number* and the *yellow locator card*.

There are, therefore, three types of "numbers" controlling the input of data:

(1) the *cross reference number;*
(2) the *précis number;*
(3) the *special research code number,* which is also included in the *cross reference number.*

ILLUSTRATION OF THE SYSTEM

Figure 2 illustrates the entry for Marquette County. The *cross reference number* in the first line is explained as follows:

125.1 - United States of America; 126.26 - State of Michigan; 127.103 - Marquette County. Entries to the right identify data from the *yellow locator card*. Within the next numerical entry on the left margin, A0138 is the précis number for Marquette County. A01382 is the data source, in this case, *Roy L. Dodge, Michigan Ghost Towns, Vol. III, Upper Peninsula*. The following 34 lines, A01384, relate the historical précis.

```
125.1 126.26 127.103              (1)MARQUETTE COUNTY(2)US GEOL SURVEY
125.1 126.26 127.103              (3)46 30'-41' 88 00-07'(6)COUNTY
125.1 126.26 127.103              (1)MARQUETTE(6)COUNTY(7)56,154
A01382      DODGE
A01384            MARQUETTE COUNTY WAS NAMED IN HONOR OF FATHER MARQUETTE, THE
A01384      JESUIT MISSIONARY. THE COUNTY WAS LAID OUT IN 1843 AND ATTACHED TO
A01384      HOUGHTON COUNTY UNTIL 1846. IT WAS REORGANIZED IN 1848, AND THE COUNTY
A01384      GOVERNMENT ORGANIZED IN 1851. THE CITY OF MARQUETTE IS THE COUNTY
A01384      SEAT.
A01384            AS ARE MOST UPPER PENINSULA COUNTIES, MARQUETTE COUNTY IS MADE UP
A01384      OF MANY NATIONAL ORIGINS, INCLUDING FRENCH, FINNISH, IRISH, CANADIAN,
A01384      GERMAN, ITALIAN, POLISH, SWEDISH AND WELSH.
A01384            EARLY HISTORY HAS BEEN AMPLY RECORDED, AND THE PETER WHITE PUBLIC
A01384      LIBRARY, IN MARQUETTE, IS REGARDED AS THE FINEST LIBRARY IN THE UPPER
A01384      PENINSULA, WITH MORE THAN 100,000 VOLUMES. THE J.M. LONGYEAR RESEARCH
A01384      LIBRARY OF THE MARQUETTE COUNTY HISTORICAL SOCIETY IS ALSO THE FINEST
A01384      OF THIS TYPE ON LAKE SUPERIOR.
A01384            THE MINING OF IRON ORE HAS ALWAYS BEEN THE MAIN INDUSTRY IN THE
A01384      COUNTY, WITH LUMBERING SECOND, BUT PRODUCTION DECLINED IN THE 1960'S
A01384      AND MANY PEOPLE WERE UNEMPLOYED. TODAY MORE MODERN METHODS OF MINING
A01384      HAVE BEEN DEVELOPED AND, WITH THE APPROVAL OF THE TILDEN MINE PROJECT,
A01384      WHICH INCLUDES THE CONSTRUCTION OF A DAM ABOUT 6 MILES UP THE MIDDLE
A01384      CALLED THE GREENWOOD RESERVOIR, FOR USE IN PROCESSING LOW-GRADE IRON
A01384      ORE, THE ECONOMY OF MARQUETTE COUNTY IS EXPECTING A BOOM UNHERALDED IN
A01384      ITS HISTORY FROM BOTH THE $200 MILLION CONSTRUCTION COST OF THE
A01384      PROJECT, WITH A $38 MILLION PAYROLL, PLUS THE NEW PELLETIZING PLANT OF
A01384      THE COMPANY, AND THE BENEFIT TO THE TOURIST INDUSTRY RESULTING FROM
A01384      THE ATTRACTION TO THE HUGE RESERVOIR.
A01384            MARQUETTE COUNTY IS ONE OF ONLY 3 UPPER PENINSULA COUNTIES WITH
A01384      AN INCREASE IN POPULATION, INSTEAD OF A DECREASE, IN THE PAST DECADE,
A01384      1960-70. THE POPULATION INCREASED DURING THIS PERIOD FROM 56,154 IN
A01384      1960 TO 64,686 IN 1970.
A01384            IN 1867, THE MICHIGAN SUPREME COURT RULED AGAINST THE LEGISLATURE
A01384      AND IN FAVOR OF THE CITIZENS OF MARQUETTE COUNTY AFTER SOME OF THE
A01384      MOST VALUABLE MINING PROPERTY IN THE COUNTY AND IN THE CITY OF
A01384      ISHPEMING HAD BEEN PLACED IN A NEW COUNTY, NAMED "WASHINGTON". THE
A01384      ORIGINAL MARQUETTE BOUNDARIES WERE RE-ESTABLISHED AND MICHIGAN LOST
A01384      WASHINGTON AS A COUNTY NAME.
A01381      125.1 126.26 127.103(009)
A01392      RANKIN
A01394      FOR FATHER JACQUES MARQUETTE, EARLY FRENCH MISSIONARY. COUNTY
A01394      ORGANIZED IN 1843. LARGEST COUNTY IN U.P. MAR-KET' COUNTY SEAT,
A01394      MARQUETTE. MARQUETTE COUNTY GOVERNMENT ORGANIZED NOV. 4, 1851.
A01394      "MARQUETTE," FR.= A CAKE OF WAX. IT CAN BE ASSUMED THAT THE SURNAME,
A01394      "MARQUETTE," CAME FROM A FAMILY OF WAX MAKERS.
A01391      125.1 126.26 127.103
125.1 126.26 127.103.01           (1)CHAMPION(2)US GEOL SURV(3)87 50'-47',
125.1 126.26 127.103.01           46 37'(4)48N 28W 19SE, 20SWSE, 21SWSE, 22SW
125.1 126.26 127.103.01           (5)1583-1453(6)TOWNSHIP(7)512
125.1 126.26 127.103.03           (1)CHOCOLAY(2)US GEOL SURV(3)87 21', 46 30'
125.1 126.26 127.103.03           (4)47N 24W 6NW(5)800(6)TOWNSHIP(7)3299
125.1 126.26 127.103.05           (1)ELY(2)US GEOL SURV(3)87 49'-47', 46 24'
125.1 126.26 127.103.05           (4)46N 28W 21NE, 22NENW(5)1450(6)TOWNSHIP
```

Figure 2 PNSUS - Upper Peninsula of Michigan: computer printout for Marquette County

The précis ends with the repetition of the *cross reference number* 125.1 126.26 127.103 with the addition of (009), noting a special research code. The following line with the number beginning A0139, indicates that a second précis card on historical information about Marquette County has been filed. The complete number, A01392, shows that the data card is an excerpt from the Ernest Rankin card file in the Marquette County Historical Society records. As in the case of précis A0138, the five lines of Rankin's précis are ended by repeating the précis A0139 plus 1 and giving the *cross reference number* for Marquette County, 125.1 126.26 127.103. Thus each précis on historical data, as well as linguistic and folkloric data, is being collated and tied in to the geographical coordinates and *cross reference number*. The next three entries for Champion, Chocolay and Ely show only yellow locator card information with no additional research data.

CONCLUSION

Goals set forth in Washington in 1978 gave direction and uniformity to the objectives of the Place Name Survey of the United States. In keeping with these aims, the PNSUS for the Upper Peninsula of Michigan has now gathered *informational* (i.e., identifying) and *explanational* (i.e., onomastic) data for the computerized bank. The present listing of 11 800 names relate to both "God-made" and "man-made" features within and around the shores of the Upper Peninsula of Michigan.

PANEL DISCUSSION AND COMMENTARY
ÉCHANGE EN PANEL ET COMMENTAIRE

★★★

Moderator/Président:

Richard Groot	Director, Geographical Services, EMR/ Directeur, Service de géographie, EMR

Panelists/Participants du forum:

Tim Evangelatos	Chief, Automated Cartography, Canadian Hydrographic Service, Fisheries and Oceans/Chef, Cartographie automatisée, Service hydrographique du Canada, Pêches et Océans
James Linders	Chairman, Department of Computing and Information Science, University of Guelph
Michael E.H. Young	Chief, Systems Engineering, Topographical Survey, EMR/Chef, Assistance technique, Division topographique, EMR
Donald Orth	Chief, Branch of Geographic Names, National Mapping Division, U.S. Geological Survey and Executive Secretary, Domestic Names Section, U.S. Geographic Board
Christian Bonnelly	Responsable du service de l'analyse et du contrôle, Commission de toponymie du Québec, Ministère du Conseil Exécutif

Audience participants/Participants de l'auditoire:

Rolf Böhme	Scientific Director, Institute of Applied Geodesy, Frankfurt am Main, Federal Republic of Germany/Directeur scientifique, Institut des levés géodésiques appliqués, Frankfurt am Main, République fédérale d'Allemagne (Institut für Angewandte Geodäsie, Frankfurt am Main)
Simon Ommanney	Head, Perennial Snow & Ice Section, National Hydrology Research Institute, Department of Environment/Chef, Section de neiges et de glaces éternelles, Institut National de Recherche en Hydrologie, Département de l'Environnement
Michael Smart	Manager, Nomenclature Section, Surveys and Mapping Branch, Ministry of Natural Resources, Ontario
Françoise Ricour-Singh	Chief, Spatial Delineation and Analysis Section, Statistics Canada/Chef, Section de la délinéation et de l'analyse, Statistique Canada
Bertrand Rivard	Directeur adjoint, Centre de recherche en aménagement régional, Université de Sherbrooke

Mr. *Groot* opened the discussion by observing that the papers concerning the creation or use of toponymic data base, presented earlier in the day, revealed the following salient points:

(1) Unanimity of opinion is that computerization is an inevitable step in the management of an ever-increasing volume of toponymic data.

(2) Many organizations, having made the decision to automate names information, are in various stages of developing, loading or partially using the data. No one single system, however, is yet fully operational.

With these ideas in mind, Mr. *Groot* directed the panel discussion towards three questions:

(1) What are users' expectations of geographical names information, which in the very near future will be available in digital form?

M. *Groot* ouvre la discussion en faisant observer que les documents présentés plus tôt au cours de la journée au sujet de la création ou de l'utilisation de bases de données toponymiques font ressortir les points suivants:

(1) Tout le monde s'entend pour dire que l'automatisation constitue une étape inévitable dans la gestion d'un volume toujours croissant de données toponymiques.

(2) Parmi les nombreuses organisations qui ont pris la décision d'automatiser leur information toponymique, certaines en sont au stage de l'élaboration ou de l'emmagasinage, et d'autres utilisent déjà partiellement leurs données. Aucun système, toutefois, n'est encore entièrement opérationnel.

Cela posé, M. *Groot* oriente le débat vers trois questions pertinentes:

(1) Qu'est ce que les usagers attendent de l'information sur les noms géographiques, qui sera bientôt disponible sous forme numérique?

(2) What should be the limits of the content of a toponymic data base? Should information relate purely to the name (i.e. origin, date of approval, coordinates, etc.) or should other information, such as population data, or length of rivers be included?

(3) How do we assure a degree of compatibility between federal and provincial systems so that the flow of information important for the two levels of operation is facilitated?

(1) USERS' EXPECTATIONS OF A TOPONYMIC DATA BASE

As representatives of two major users, Mr. Young and Mr. Evangelatos were asked the views of Topographical Survey, EMR and of Canadian Hydrographic Service (CHS) respectively.

Mr. *Young* explained that Topographical Survey has now developed and implemented a digital mapping system for 1:50 000 scale sheets. Unlike the previous experimental system in which data were digitized from a graphic source, data are now collected directly from aerial survey source information. So far placement of names has remained a manual operation, in part due to technological development priority being given to topographical data, and in part due to the vast number of cartographic judgments that would need to be made. At present the technological problems in providing a draftsman with an interactive system include the lack of large enough display devices and a slow response time, whereby manual type placement is still a quicker operation. As yet all type fonts needed cannot be drawn automatically to a

(2) Quelles devraient-être les limites du contenu d'une base de données toponymiques? L'information devrait-elle porter uniquement sur le nom (c'est-à-dire sur son origine, la date de son approbation, ses coordonnées, etc.), ou devrait-elle englober d'autres informations, comme des données démographiques ou, par exemple, la longueur des rivières?

(3) Comment établir une certaine compatibilité entre les systèmes fédéraux et provinciaux, de manière à faciliter la circulation de l'information importante pour les deux ordres de gouvernement?

(1) QU'ATTENDENT LES USAGERS D'UNE BASE DE DONNÉES TOPONYMIQUES

On demande à MM. Young et Evangelatos les opinions des deux usagers importants qu'ils représentent, à savoir le Service des levés topographiques, EMR, et le Service hydrographique du Canada (SHC).

M. *Young* explique que le Service des levés topographiques vient de créer et de mettre en oeuvre un système de cartographie numérique pour des feuilles à 1/50 000. Contrairement au système expérimental précédent, dans lequel on numérisait des données graphiques, les données sont maintenant puisées directement dans des informations produites par des levés aériens. Jusqu'ici, le placement des noms est demeuré une opération manuelle, en partie parce que les efforts de développement technologique ont été concentrés sur les données topographiques, et en partie parce qu'il faut porter un grand nombre de jugements d'ordre cartographique. Actuellement, l'emploi d'un dessinateur équipé d'un système interactif pose des problèmes d'ordre technologique; par exemple, on ne dispose pas d'appareils d'affichage suffisamment grands et les temps de réponse sont longs.

Photo: Mike Kelly, Information EMR

SYMPOSIUM PANELISTS/PARTICIPANTS DU FORUM:

(Left to right/de gauche à droite): Christian Bonnelly, Donald Orth, Richard Groot (Moderator/Président), Tim Evangelatos, Michael Young, James Linders quality sufficiently high for reproduction. A method of flashing symbols from a photohead on an automatic plotter could be used, but current restrictions include the limited sizes of photoheads available and the limited number of characters on one photodisk. A cathode-ray type of photohead in which characters are formed and imaged on the drafting medium is now a possible, although expensive, alternative. If placement of geographical names is to be automated, Topographical Survey would require the ability to call up names on a map-sheet area basis, and would need the information (e.g., generic, population figures) necessary to decide on type size and font. The data should include locational reference at a high enough resolution to identify point features unambiguously at any particular scale, and to identify the extent of linear and areal features.

Mr. *Evangelatos* agreed with many of the reasons for delaying automating names placement in automated cartographic systems, but he pointed out that CHS is using a plotter which is far less expensive than cathode-ray photoheads and which gives slow, but satisfactory, results. CHS requirements, however, are more modest than Topographical Survey's, including a catalogue of approximately 5000 names for underwater features, plus a minimal quantity of names for topographical features. Storage in digital form facilitates production of gazetteer-style listings, but the development of the names editing capabilities of the systems are only just becoming a priority and necessary software will be purchased. For CHS it would be most efficient to obtain names data from a specialized digital data base, thus ensuring up-to-date, correct nomenclature in both English and French. For names of topographical features there would be a need for the geographic position and enough supporting data to determine whether a name should appear on a chart of a particular scale. Like Topographical Survey, CHS would need to be able to identify type size and font.

Mr. *Böhme* was asked to speak on developments at the Institute of Applied Geodesy as far as the user, especially in the cartographic field, is concerned in designing the capabilities of a toponymic data base. He responded that the concise Gazetteer of the Federal Republic of Germany, to be published in 1981, was being produced to fulfill the requirements of a UN resolution. Names were limited to those shown on 1:500 000 scale maps; even so all the work of automated data capture has so far taken 12 person-years.

Dr. *Linders,* from the viewpoint of a computer scientist, saw considerable difficulties in completely automating names placement, particularly in the handling of the user's spatial conception.

(2) LIMITS TO THE CONTENT OF A TOPONYMIC DATA BASE

Dr. *Linders* emphasized that the success of a toponymic data base would depend on it being kept as a simple file system, and not building into it other file records and the complex structure of a geo-reference data base or geographical information system. Such items as connectivity of a river system or special relationships (e.g. rivers and soils) he suggested were beyond the scope of a toponymic data base.

Mr. *Orth* indicated that the primary goal of the U.S. Geographic Board is to standardize the written form of a geographic name and to identify the extent of its application. However, once geographic coordinates are applied to a feature it was found that users other than mapping agencies (e.g. shipping companies) become interested in the data file.

C'est pourquoi une opération manuelle est encore ce qu'il y a de plus rapide. De plus, on ne peut tracer automatiquement toutes les fontes de caractères avec une qualité suffisante pour la reproduction. On pourrait utiliser une méthode de flash dans la tête photographique d'un traceur automatique, mais cette solution est présentement limitée par la taille des têtes disponibles et par le nombre de caractères entrant sur un disque photo. Une solution possible, quoiqu'onéreuse, serait l'emploi d'une tête à rayons cathodiques, au moyen de laquelle les caractères seraient formés et reproduits sur le support. Si la mise en place des noms géographiques devait s'effectuer automatiquement, le Service des levés topographiques aurait besoin de techniques capables d'appeler les toponymes sur une feuille de carte, ainsi que des informations (par exemple des génériques, des données démographiques) nécessaires pour déterminer le type de caractères et leurs dimensions. Ces informations devraient comprendre des coordonnées suffisamment précises pour localiser des détails ponctuels sans ambiguïté quelle que soit l'échelle, et identifier l'étendue des détails linéaires et zonaux.

M. *Evangelatos* accepte nombre des raisons invoquées pour retarder l'automatisation de la mise en place des noms dans les systèmes de cartographie automatique, mais il signale que le SHC emploie actuellement un traçeur qui coûte beaucoup moins cher que les têtes photographiques à rayons cathodiques et qui peut produire des résultats peu rapides, mais satisfaisants. Toutefois, les besoins du SHC sont plus modestes que ceux du Service des levés topographiques; ils se réduisent à un catalogue d'environ 5000 noms désignant des détails sous-marins, plus un petit nombre de noms d'accidents topographiques. L'emmagasinage sous forme numérique facilite la production de listes genre répertoires de noms géographiques, mais la mise au point de techniques d'édition ne fait que commencer à devenir une priorité, et le logiciel nécessaire sera acheté. Pour le SHC, le mieux serait de puiser les informations toponymiques dans une base de données numériques spécialisée, ce qui lui procurerait une nomenclature précise et à jour en anglais et en français. En ce qui concerne les détails topographiques, le SHC aurait besoin des coordonnées géographiques et de suffisamment de données complémentaires pour déterminer si un nom devrait figurer sur une carte à telle ou telle échelle; de plus, à l'instar du Service des levés topographiques, il devrait disposer de l'information nécessaire pour choisir le type de caractères et leurs dimensions.

On demande à M. *Böhme* de parler des progrès réalisés à l'Institut de géodésie appliquée dans la création d'une base de données toponymiques, susceptibles d'intéresser l'utilisateur, en particulier dans le domaine cartographique. M. *Böhme* répond que le répertoire toponymique concis de la République fédérale d'Allemagne, qui doit paraître en 1981, est produit à la suite d'une résolution des Nations unies. Bien que le répertoire se limite aux noms qui apparaissent sur les cartes à 1/500 000, tout le travail de la saisie automatique des données a nécessité 12 années-personnes jusqu'à maintenant.

En tant qu'informaticien, M. *Linders* entrevoit beaucoup de difficultés à automatiser le placement des écritures, surtout en tenant compte de la conception spatiale de l'utilisateur.

(2) LIMITES DU CONTENU D'UNE BASE DE DONNÉES TOPONYMIQUES

M. *Linders* souligne qu'une base de données toponymiques sera réalisée avec succès dans la mesure où elle demeurera un système de fichiers simple, et que l'on ne tentera pas d'y intégrer d'autres enregistrements de fichiers, ni de lui attribuer la structure complexe d'une base de géo-références

Photo: Mike Kelly, Information EMR

AUDIENCE PARTICIPANT/PARTICIPANT DE L'AUDITOIRE: Rolf Böhme

With horizons expanded for the production of gazetteers and answers to inquiries, plans exist to extend the information to cover items such as population data and elevations.

Dr. *Linders* expressed concern for data base consistency if data do not reside with their mandate area. He agreed that a geographic position is a meaningful geo-code for correlating files, but pointed out that a data base becomes quickly corrupted if extended to cover another mandate area. It would be preferable to have a utility environment in which each application area maintains its own files, but has access to the other file systems to answer questions of common concern. Toponymy would appear to be a very good application area.

Mr. *Orth* confirmed that the U.S. Geographic Board was very aware of maintaining the integrity and security of a data base, and was only adding elements that had direct relevance to the geographic names data base.

Mr. *Bonnelly* added that in Quebec, each department, in carrying out its specific mandate, follows standards to collect technical data, such as mountain elevations and population of communities. The mandate of the Commission de toponymie is to provide accurate names. The Commission hopes to have access to the information systems of various departments, to obtain the name for an entity, together with various other non-toponymic parameters.

Mr. *Böhme* emphasized the necessity for generalization in cartography, and the need for information to decide on size and weight of symbols and type. To meet these requirements the Institute saw a need to include in its data base items, such as population, length of rivers, area of lakes, presence of facilities (e.g. airports or railway stations).

ou d'un système d'information géographique. Selon lui, des éléments d'information tels que la délimitation d'un réseau fluvial ou des rapports spatiaux (par exemple, les cours d'eau et les sols) débordent le cadre d'une base de données toponymiques.

M. *Orth* indique que le but premier du U.S. Geographic Board est de normaliser l'orthographe d'un nom géographique et d'en déterminer le domaine d'extension. Cependant, on constate qu'une fois des coordonnées géographiques attribuées à un détail, des utilisateurs autres que les organismes de cartographie (des sociétés d'expédition, par exemple) commencent à s'intéresser au fichier de données. Maintenant que le Geographic Board s'occupe de la production de répertoires et répond à des demandes de renseignements, on envisage d'étendre l'information aux données démographiques et aux altitudes.

M. *Linders* fait valoir que la cohérence d'une base de données est menacée si les données qu'elle renferme ne relèvent pas de la responsabilité de l'organisme qui en a la charge. Il est d'accord pour dire qu'une position géographique représente un géo-code utile pour mettre des fichiers en corrélation, mais il ajoute qu'une base de données dégénère rapidement si on l'étend à d'autres domaines. Il serait préférable, selon lui, qu'il se crée un environnement utilitaire dans lequel chaque domaine d'application aurait ses propres fichiers, tout en ayant accès aux autres systèmes de fichiers pour répondre à des questions d'intérêt commun. Il semblerait que la toponymie soit un très bon domaine d'application.

M. *Orth* rassure ses interlocuteurs en précisant que le U.S. Geographic Board est très soucieux de maintenir l'intégrité et la sécurité d'une base de données, et qu'il ne fait qu'ajouter des éléments qui ont un rapport direct avec la base de données sur les noms géographiques.

M. *Bonnelly* ajoute qu'au Québec chaque ministère poursuit des travaux bien définis et suit des normes pour recueillir des données techniques telles que l'élévation des montagnes et la population des communautés de villages. Le mandat de la Commission de toponymie est de fournir des noms précis. Ce que la Commission veut c'est de pouvoir rendre les systèmes informatiques des différents ministères accessibles pour avoir un nom pour une entité accompagné par différents paramètres autres que les paramètres toponymiques.

M. *Böhme* met en relief la nécessité d'une généralisation en cartographie, ainsi que le besoin d'informations permettant de choisir la taille et l'importance des symboles et des caractères. Pour satisfaire à ces exigences, l'Institut entrevoit la nécessité d'inclure dans sa base de données des éléments d'information tels que la population, la longueur des cours d'eau, l'étendue des lacs, les installations existantes (comme des aéroports ou des gares). Afin de faciliter la production de cartes thématiques, on créera, pour ce genre de données, des fichiers dont les entrées seront associées à des numéros de cartes topographiques.

M. *Groot* souligne que, bien que la création d'une base de données pour le récent Atlas toponymique du Canada ait semblé facile, de nombreux problèmes ont été posés par la fusion des données de recensement de Statistique Canada et de l'information du CPCNG au sujet des noms des endroits habités.

M. *Linders* a insisté sur les différences entre fonction toponymique et fonction cartographique, et sur la nécessité de dissocier ces deux domaines.

On demande à M. *Ommanney* de donner, à titre de glaciologue et de compilateur de définitions de termes génériques, son opinion sur ce que devraient être les limites de la base de

To aid production of thematic maps, files are being set up for this type of data, and entries will be linked to topographic map numbers.

Mr. *Groot* pointed out that, although it sounded easy, many problems had arisen in merging the records of Statistics Canada census data and CPCGN populated place name information to create a data base for the recently published Canada Gazetteer Atlas.

Dr. *Linders* stressed that toponymic and cartographic functions were different and should be kept distinct.

As a glaciologist and compiler of definitions of generic terms, Mr. *Ommanney* was asked for his views on the extent of the data base. He thought that much stress in toponymy was placed on point locations, whereas in fact ice and marine features, for example, have very definite spatial characteristics. The delineation of such features he felt to be an integral part of the whole naming process. He did not think that the extent of application of mountain or glacier names should fall to the cartographer, but should lie within the mandate of those developing a toponymic data base.

According to Mr. *Bonnelly*, the Commission's staff often has problems when it comes, for example, to identifying the main branch or the source of a river. The aid of hydrology experts is sought to correctly assess the geographical context in which a given name is applied.

Mr. *Evangelatos* showed that similar hydrographic problems could only be solved for the cartographer by consultion between the field hydrographer, nomenclature staff and, if necessary, local residents. Mr. *Young* felt that it should be the responsibility of the topographer to define the extent of geographical features from interpretation of photographic imagery and field data. However, as Mr. *Smart* pointed out the limits of a geographical feature interpreted from a photo or map may be very different from the conceptual perspective of a name application according to local usage. *Lac Deschênes* is a good example of a case where field interpretation of the name application differed (and took precedence) over cartographic interpretation of the extent of the feature. Mr. *Groot* supported the importance of local toponymic use, and underlined this as a definite problem in automating geographical names data.

Mr. *Ommanney* asked how cartographers, if using automated names placement, deal with a feature that, because of its areal extent, should be named on several adjoining map sheets. If names are pulled from the data base in accordance with their geographical coordinates they may be omitted from names listings for some map sheets.

Mr. *Orth* explained that the U.S. Geographic Board had interpreted existing topographic maps to collect information. Any name could, if applicable, then be tied in to several map sheets. In the future, line and point data could be digitized for the cartographer, but present users must still rely on existing maps for much information concerning the extent of features. In the introduction to a state gazetteer there will be included an indication of the degree of accuracy in coordinate designation, and an explanation of how coordinates, for such locations as river mouths or bays, have been determined. Mouths and heads of rivers have been identified to the nearest second. If on perusal, experts show corrections to be necessary, these can be accommodated.

As far as Mr. *Young* was concerned he did not feel pinpointing the exact location of a river mouth to be of par-

données. Selon lui, les toponymistes attachent beaucoup d'importance aux détails ponctuels, alors qu'en fait, la glace et les phénomènes marins, par exemple, présentent des caractéristiques spatiales bien définies. Il est d'avis que la délimitation de ces éléments doit faire partie intégrante de l'ensemble du processus d'appellation. D'après lui, le domaine d'extension des noms de montagnes ou de glaciers ne devrait pas intéresser le cartographe, mais s'inscrire dans le mandat des personnes chargées de la création d'une base de données toponymiques.

M. *Bonnelly* dit qu'à la Commission de toponymie ils ont souvent un problème quand il s'agit d'identifier la branche principale ou la source d'un cours d'eau. C'est dans un cas semblable qu'ils ont besoin de la contribution de spécialistes en hydrologie pour pouvoir préciser correctement le contexte géographique où ils doivent appliquer correctement le nom.

M. *Evangelatos* indique que des problèmes hydrographiques semblables qui se posent au cartographe ne peuvent être résolus que par consultation entre l'hydrographe en place, le personnel préposé à la nomenclature et, au besoin, les résidents de la région. M. *Young* est d'avis qu'il devrait appartenir au topographe de définir l'étendue des entités géographiques à partir de l'interprétation de photographies et de données recueillies sur le terrain. Toutefois, M. *Smart* signale que les limites d'un détail géographique obtenues par interprétation d'une photo ou d'une carte peuvent différer sensiblement de l'idée que se font les résidents de la région de l'application des noms. Le *Lac Deschênes* constitue en celà un bon exemple; en effet, il s'agit d'un cas où l'interprétation locale de l'application du nom a préséance sur l'interprétation cartographique de l'étendue de la nappe d'eau. M. *Groot* souligne lui aussi l'importance de l'usage local des noms géographiques, la présentant comme un problème précis de l'automatisation des données toponymiques.

M. *Ommanney* demande ce que feraient des cartographes qui utilisent la mise en place automatique des noms si, en raison de son étendue, un élément devait être nommé sur plusieurs feuilles de carte contiguës. Si l'on se sert des coordonnées géographiques pour appeler les noms sur la carte, il se pourrait bien que des toponymes soient oubliés sur certaines feuilles de carte.

M. *Orth* explique que le U.S. Geographic Board a interprété des cartes topographiques déjà établies pour recueillir des informations. N'importe quel nom pourrait alors, le cas échéant, être rattaché à plusieurs feuilles de carte. Il est possible que, dans l'avenir, les données linéaires et ponctuelles puissent être converties en numériques à l'intention du cartographe; pour l'instant toutefois, les utilisateurs doivent encore compter sur des cartes déjà établies pour recueillir une bonne partie de l'information dont ils ont besoin au sujet de l'étendue des détails. Dans l'introduction du répertoire géographique d'un état, on indiquera le degré de précision des coordonnées et on expliquera comment ont été déterminées les coordonnées des endroits tels que les embouchures de rivières ou les baies. L'emplacement des embouchures et des sources de cours d'eau a été établi à la seconde près. Si, à la lecture, des spécialistes constatent qu'il y a des corrections à apporter, il est alors possible de le faire.

De l'avis de M. *Young*, la détermination de l'emplacement exact de l'embouchure d'un cours d'eau n'est pas particulièrement importante pour le cartographe, car celà ne pose ticular concern to the cartographer, as no problem of name placement was involved, and a user could draw his own conclusions about the position.

Dr. *Linders* felt there was a lack of consistency in outlook here. A point, such as a river mouth, is not reproduceable cartographically, resulting in a user not being able to determine the precise coordinates in order to access the name, and hence obtain other relevant information in the toponymic data file. Mr. *Orth* suggested that if a river mouth location was so ambiguous (e.g. had many distributaries) and the user still needed to access by location, sufficient tolerance could be built into the search so that the name of the feature in question could be picked up.

Mr. *Böhme* spoke of the special hierarchical numbering system that the Institute was using for the description of components of drainage networks. A similar numbering system is being used for mountain range hierarchies. These designations are being used instead of coordinates in the current Gazetteer of the Federal Republic of Germany; they also link well to other data bases. Mr. *Orth* commented upon the hierarchical coding system being used in Australia in a cartography-related information base, to identify named features of different levels of importance for application to different scale maps. Mr. *Young* agreed that the concept of a cartographic data base is a hierarchical one containing formatted information to locate, classify and sub-classify features and unformatted (textual - e.g. names) information. Ideally an extensive Canadian cartographic data file will be developed, in which the name is a unit of information that can be related back to the toponymic file if more detail is required for it.

In summarizing this particular question Mr. *Groot* felt a consensus was heard that a toponymic data base should remain toponymic. Limits of application of a name to any particular feature should be included, but extension to include data outside the toponymic field requires considerable caution. In particular, such steps could lead to storage of ambiguous or out-of-date data and production of a data base management mode considerably more complex to implement and resolve than a file retrieval system.

(3) COMPATIBILITY BETWEEN FEDERAL AND PROVINCIAL SYSTEMS

Mr. *Groot* requested comments concerning custodians of toponymic information at the federal and provincial levels and how the functions of each authority should be fulfilled with free flow of information from one to another.

Mr. *Bonnelly* noted that, elements in the Commission's data bank could be represented by a parameter which would designate features that should be identified on hydrographic charts or on maps at the 1:25 000, 1:50 000, 1:250 000 and 1:500 000 scales. If the systems of the Commission and the federal Surveys and Mapping Branch were reciprocal, the results would be very positive. The Commission encounters technical compatibility difficulties, where hardware and instrumentation are concerned, with computerized systems of other departments wishing to use their services. However, the basic elements and toponymic data will easily be interchangeable within the specific framework of each department as established by law.

aucun problème de mise en place des noms et un utilisateur peut tirer ses propres conclusions au sujet de la position de l'élément considéré.

M. *Linders* signale une incohérence. Un point tel que l'embouchure d'un cours d'eau n'est pas cartographiquement reproductible, de sorte qu'un utilisateur est incapable d'en déterminer les coordonnées exactes pour avoir accès au nom et, par le fait même, d'obtenir d'autres informations pertinentes dans le fichier des données toponymiques. M. *Orth* suggère que, si l'emplacement de l'embouchure d'un cours d'eau est à ce point ambigu (par exemple, si la rivière a de nombreux défluents) et que l'utilisateur ait encore besoin de trouver l'endroit par ses coordonnées, il est possible de prévoir dans la recherche une tolérance suffisante pour pouvoir trouver le nom de l'élément considéré.

M. *Böhme* parle du système particulier de numérotation hiérarchique que l'Institut emploie pour décrire les éléments constitutifs d'un réseau hydrographique. Un système analogue est utilisé pour hiérarchiser les chaînes de montagnes. Ces désignations sont substituées aux coordonnées dans l'actuelle version du répertoire de la République fédérale d'Allemagne, et rattachées à d'autres bases de données. M. *Orth* parle à son tour du code hiérarchique que l'Australie emploie dans une base de données cartographiques pour attribuer aux éléments divers niveaux d'importance, selon l'échelle des cartes où on les utilise. M. *Young* est d'accord pour dire que le concept de base de données cartographiques sous-tend une idée de hiérarchie et qu'il englobe à la fois des informations structurales permettant de situer, de classer et de subdiviser des éléments, de même que des informations non structurales, comme des toponymes. Idéalement, on créera au Canada un vaste fichier de données cartographiques, dans lequel le nom sera une unité d'information qui pourra nous renvoyer au fichier toponymique s'il nous faut davantage de renseignements à son sujet.

Pour résumer le débat sur cette question, M. *Groot* dit qu'il croit avoir dégagé un consensus: une base de données toponymiques devrait demeurer toponymique. Il faudrait y inclure le domaine d'extension d'un élément donné, mais l'introduction de données étrangères au domaine toponymique exige beaucoup de prudence. Pareille pratique pourrait notamment aboutir à l'emmagasinage de données ambiguës ou périmées et à un mode de gestion de données beaucoup plus difficile à appliquer qu'un système de consultation de fichiers.

(3) COMPATIBILITÉ ENTRE LES SYSTÈMES FÉDÉRAUX ET PROVINCIAUX

M. *Groot* demande des renseignements au sujet de la garde de l'information toponymique au sein des gouvernements fédéraux et provinciaux et de la manière dont chaque autorité devrait exercer ses fonctions de façon que l'information circule librement d'un organisme à un autre.

M. *Bonnelly* fit remarquer que les éléments de leur banque de données pourraient être identifiés par un paramètre qui pourrait décrire les éléments qui devraient paraître sur les cartes hydrographiques ou les cartes au 25 000, 50 000, 250 000 et 500 000. À ce niveau là, si le système de la Commission et celui de la Direction des levés et de la cartographie au fédéral pourraient être compatibles, ce serait très positif. La Commission rencontre des problèmes de compatibilité technique quant à la quincaillerie et l'instrumentation avec les systèmes informatiques des autres ministères qui voudraient s'assurer nos services. Cependant, les éléments et données toponymiques de base seront facilement interchangeables à l'intérieur des cadres spécifiques prévus par les lois respectives.

Dr. *Linders* expressed concern over three aspects: (a) the necessity to establish a common nomenclature - as by common data dictionaries currently being prepared by Canadian committees, to allow intercommunication between all federal and provincial surveying agencies; (b) the referencing of entities by geo-codes - a new area of the technology currently emerging; and (c) the sharing of information - intercommunication protocols, procedures and semantics are now being developed which will allow computers to 'speak' to each other over telephone lines.

M. *Linders* se dit préoccupé par trois aspects de la question: (a) la nécessité d'établir une nomenclature commune, comme c'est le cas des dictionnaires de données actuellement préparés par des comités canadiens pour faciliter la communication entre tous les organismes fédéraux et provinciaux chargés des levés; (b) la désignation d'entités au moyen de géo-codes, nouveau domaine de la technologie; et (c) le partage de l'information (on est à mettre au point des protocoles, des méthodes et une sémantique de communication qui permettront à des ordinateurs de "se parler" par le truchement de lignes téléphoniques).

In the United States, Mr. *Orth* reported, the situation is a little different. No state has its own file on geographic names in machine readable form, but each depends on the federal body delegated by law to establish official geographical names and their applications. Communication within federal departments, agencies and state governments is the first step. The U.S. Geographic Board is coordinating the official data file, which will then be compared with other existing data files, for example those of the National Bureau of Standards and the Census Bureau.

M. *Orth* signale qu'aux États-Unis, la situation est quelque peu différente. Aucun état ne possède son propre fichier sur les noms géographiques dans une forme assimilable par l'ordinateur, mais chacun dépend de l'organisme fédéral désigné par la loi pour établir des noms géographiques officiels et leurs applications. La communication au sein des ministères fédéraux, des organismes et des gouvernements des états constitue la première étape. Le U.S. Geographic Board coordonne le fichier des données officielles, qui sera ensuite comparé à d'autres fichiers de données existants, comme ceux du National Bureau of Standards et du Census Bureau.

Dr. *Ricour-Singh* reported that Statistics Canada has a computerized geographical file listing names of all municipalities and census divisions and a secondary file containing the names of 15 000 localities. However, because of delimitation difficulties, there are no files for localities within cities or urbanized areas. The names of localities are causing the most problems, because these names are assigned in the field by the census representatives, who define what they or the local population consider to be a locality. They give the names according to what is heard in the field; sometimes these names differ from those approved by the CPCGN. Unapproved names of localities are submitted to the CPCGN for approval.

Mme *Ricour-Singh* dit qu'à Statistique Canada ils ont un fichier géographique où ils entreposent les noms de toutes les municipalités, des divisions de recensement et un fichier secondaire qui contient les noms de 15 000 localités. Cependant, ils n'ont pas de fichier pour les localités qui sont dans les villes ou les régions urbanisées à cause des difficultés de délimitation. Ce sont les noms des localités qui leur causent le plus de difficulté. C'est que les noms des localités sont définis par les agents recenseurs qui vont sur le terrain et définissent ce qu'eux voient ou ce que la population voit comme étant une localité. Ils donnent les noms qu'ils ont entendus sur le terrain et il arrive que ce nom diffèrent de deux approuvés par le Comité permanent. Les noms des localités qui ne sont pas approuvés sont soumis au Comité permanent pour son approbation.

Mr. *Bonnelly* indicated that the Commission de toponymie could assist in providing a quantitative definition of *localité,* as opposed to *hameau*. The Commission could also coordinate its research with Statistics Canada.

M. *Bonnelly* souligne que la Commission de toponymie pourrait apporter sa contribution pour fournir une définition quantitative de *localité* par rapport à celle que l'on donne à *hameau* et pourrait orienter ses recherches avec Statistique Canada.

Mr. *Rivard* observed that at the Centre de recherche en aménagement régional, space configurations are fed into a computer for use on maps produced entirely by computer. Computerized graphical representation offers many technical advantages and is more economical. For these projects to materialize, much depends on the CPCGN or the Commission de toponymie. What the Centre requires from these two services is the official designation on which to base the processing of information, whether manually or automatically, in order to interrelate various data tapes. Requirements of organizations other than the Commission and the CPCGN may then be answered. To be able to define each characteristic precisely, it would be useful if everyone involved worked together and if each discipline within its own specialty provided support.

M. *Rivard* souligne qu'au Centre de recherche en aménagement régional on transpose en ordinateur des configurations spatiales pour fin d'utilisation de cartes qui sont produites entièrement par ordinateur. Pour ce qui a trait à la représentation graphique automatisée, il y a beaucoup d'avantages techniques et c'est plus économique. Pour réaliser ces projets on attend beaucoup soit du Comité permanent ou de la Commission de toponymie. Ce qu'ils attendent de ces deux organismes c'est de leur fournir la désignation officielle sur laquelle ils pourront s'asseoir pour pouvoir traiter de l'information, soit manuellement, soit automatiquement, afin de pouvoir interrelier différentes bandes de données. Ces bandes peuvent être localisées à différents endroits et répondre aux besoins d'organismes autres que ceux de la Commission ou ceux du Comité permanent. Pour bien définir un trait, il serait bon que tout le monde travaille ensemble et que chaque discipline, dans sa spécialité apporte sa collaboration.

In summary, Mr. *Groot* noted that there is certainly a requirement for sharing the information between custodians of toponymic data bases. Problems identified in sharing these data lie in the creation of dictionaries of common language to communicate between the data banks and to address software and hardware problems. Perhaps under the auspices of the CPCGN a pilot project should be set up to address these aspects on a small scale, taking into account the work already being done in the Canadian Council on Surveying and

En résumé, M. *Groot* souligne qu'il existe certainement un besoin de partage de l'information entre les organismes chargés de tenir les bases de données toponymiques. Ce partage de données exige toutefois la création de dictionnaires dans un langage uniforme, qui permettent de mettre en communication les banques de données et de résoudre les problèmes de logiciel et de matériel. Peut-être que, sous les auspices du CPCNG, Mapping, with respect to the creation of hierarchies of information and definition of generics in the topographic field.

CONCLUSION

Mr. *Groot* expressed his pleasure at the interest shown in the symposium, and in particular with the positive response to the automated National Toponymic Data Base (NTDB) being developed in the Surveys and Mapping Branch. He was very gratified to hear the encouraging comments of people who might like to implement such a system in their own organization.

Mr. *Groot* pointed out that, although the NTDB system is working and has proven itself, it should be realized that it was conceived and designed to meet the particular requirements of the Surveys and Mapping Branch of the federal government. These requirements may not be the same as for other organizations, and thus may influence the choice of a system others may wish to implement. The Branch's experience is that every application needs a carefully designed systems requirement leading to the selection of necessary capabilities in software and hardware which, under most circumstances, are trade-offs between the possible, the practical and the economical. Mr. *Groot*, therefore, urged participants to make a carefully planned requirement study before implementing a system.

Mr. *Groot* emphasized that the state of the art in data base management and in hardware development is rapidly advancing. The choices in equipment and systems of only two years ago are not all available now. Although a study has not been made, it is conceivable that today the Branch would choose different equipment and systems.

Mr. *Groot* stressed that an important design consideration is compatibility between provincial systems and the Branch's toponymic data base to ensure a smooth and uncomplicated flow of names information. This does not mean that systems must be identical, but that in the organization and identification of the data, compatibility is created to allow the exchange of information.

pourrait-on mettre sur pied un projet pilote au moyen duquel on expérimenterait ces aspects de la question sur une petite échelle, en tenant compte du travail qu'effectue actuellement le Conseil canadien des levés et de la cartographie en ce qui concerne la création de hiérarchies d'information et la définition des génériques en topographie.

CONCLUSION

M. *Groot* se dit très heureux de l'intérêt porté au colloque et, en particulier, de la réaction positive des participants vis-à-vis de la base automatisée de données toponymiques canadiennes mise sur pied par la Direction des levés et de la cartographie. Il est content d'entendre les commentaires encourageants de gens qui aimeraient peut-être mettre en oeuvre un système de ce genre au sein de leur propre organisation.

M. *Groot* souligne que, même si le système fonctionne et a déjà fait ses preuves, il faut se rappeler qu'il a été conçu en fonction des exigences de la Direction des levés et de la cartographie du gouvernement fédéral. D'autres organisations auront peut-être des besoins différents, et elles devront en tenir compte dans le choix de leur système. Si l'on en juge par l'expérience de la Direction, il importe d'établir avec soin les exigences de chaque application et de choisir en conséquence les ressources nécessaires en logiciel et en matériel; celles-ci résultent, dans la plupart des cas, d'un compromis entre le possible, le pratique et l'économique. M. *Groot* exhorte donc les participants à entreprendre une étude sérieuse des besoins avant de mettre un système en usage.

M. *Groot* ajoute qu'il faut aussi bien comprendre que la technologie de la gestion des bases de données et du matériel évolue rapidement. Déjà, on ne trouve plus sur le marché le choix d'équipements et de systèmes qui était offert il y a deux ans à peine. Bien qu'aucune étude n'ait encore été faite, il est concevable que la Direction, si elle avait à choisir maintenant, opterait pour un équipement et des systèmes différents.

M. *Groot* insiste sur le fait que l'un des facteurs importants dont il faut tenir compte dans la conception d'un système est qu'il doit exister une compatibilité entre les systèmes provinciaux et la base de données toponymiques de la Direction, afin d'assurer une circulation simple et facile de l'information toponymique. Non pas que les systèmes doivent être identiques, mais, dans l'organisation et l'identification des données, il faut faire en sorte que l'on puisse échanger des renseignements.

★★★★★★★★★★★★★★★★★★★★

SYMPOSIUM ON THE AUTOMATION OF
GEOGRAPHICAL NAMES

COLLOQUE SUR L'AUTOMATISATION DES
NOMS GÉOGRAPHIQUES

★

June 4, 1980 | Ottawa | le 4 juin 1980