AN INTERDEPARTMENTAL APPLICATION OF SPATIAL DATA BASES-BUILDING A STATISTICAL NETWORK FILE FROM A TOPOGRAPHIC FEATURE FILE

Joel Z. Yan
Rennie Molnar
Jean-Pierre Parker
Statistics Canada
Ottawa, K1A 0T6, Canada
J. Glen Gibbons
Energy, Mines and Resources Canada
Ottawa, K1A 0E9

BIOGRAPHICAL SKETCHES

Joel Yan is currently Chief of Methodology in Geocartographics Subdivision of Statistics Canada where he has been employed as a specialist in geographic information processing and system development for ten years. Rennie Molnar is Head of the Software Development Unit in the Geocartographics Subdivision. Geocartographics is a service bureau to Statistics Canada and other agencies in computer-assisted geography, cartography and graphics. Jean-Pierre Parker is Head of Basefile in the Geocoding Unit of Geography Division. The Geocoding Unit is responsible for creation and maintenance of spatial base files including the Area Master Files for Statistics Canada. Glen Gibbons is currently Scientific Advisor to the Director General of Surveys and Mapping Branch of EMR. Glen has eight years of experience developing digital mapping applications at EMR.

ABSTRACT

For 15 years, Statistics Canada has been building and maintaining digital street network files as a base for micro-area retrieval of census data. The total coverage now includes approximately 150 municipalities in 41 metropolitan areas containing more than half of the Canadian population. This base file is also used throughout Canada by municipal government and emergency dispatch agencies. In order to expand coverage of these files while increasing their application and compatibility with traditional map bases, digital data exchange experiments have recently been undertaken with Surveys and Mapping Branch of Energy, Mines and Resources nada (EMR) and other agencies. This paper describes the result of one experiment which involves building the network Area Master File for the City of Lethbridge, Alberta from a digital topographic file. The topographic file had been extracted from the National Topographic Data Base which is being created for a variety of uses, the primary application being the production of the 1:50,000 NTS map series. This

feature file was converted to network form at EMR, and then attribute information such as street names and addresses was added at Statistics Canada. The resulting process, scheduled to be complete by March 1985, holds promise of building a composite digital file which is both compatible with topographic maps and also has potential applications to transportation planning. Furthermore, it is hoped that AMF production costs will be reduced to permit an expansion of the AMF programme for Canada. The potential benefits of this type of standardization as well as related problems of transfer and manipulation of spatial data bases are discussed in this paper.

1. INTRODUCTION

This paper first examines the digital base file and mapping programs at the two agencies involved, and events leading up to the decision to conduct an inter-departmental pilot test using data for the City of Lethbridge, Alberta. Section two looks at the methods used and development required in conducting the test. Finally, Section three looks at the results, provides guidelines for future interdepartmental spatial data base work and raises a number of questions.

1.1 The Area Master File

Statistics Canada initiated the creation of digital street network files prior to the 1971 Census as a tool first for micro-retrieval of census data by user specified area, and later for automated collection maps (Yan and Bradley, 1983). These Area Master Files contain a logical representation of all city streets and other geographic features such as railroad tracks, rivers, and municipal boundaries in machine readable form. The AMF is maintained by the Geocoding Unit of Geography Division. It corresponds in function to the GBF/DIME file created in the U.S. during the same period, and with the TIGER file (Broome, 1984) being prepared for the 1990 Census. There are, however, differences in structure, since the DIME file is based on the block, and the AMF is based on the block-face.

Large urban areas (population 50,000 or more) are divided into block faces. A block-face consists of one side of a street between two successive intersections. These block-face spatial units are small enough that when aggregated they become a good approximation for a user identified query area. Each block face is assigned central x-y UTM coordinates, to which files of households, or persons can be coded i.e., geocoded. A user needing information from a geocoded file, outlines the area of interest on a map. This area is digitized and becomes a special "query area". All block face centroids falling within this area are aggregated and statistical data from the Census are tabulated for those block faces. This process is described in the booklet "Facts by Small Areas" (Statistics Canada, 1972). New applications of the AMF have been described by Boisvenue and Parenteau (1982). Area Master Files now exist for virtually all urbanized portions of the 38 tracted centres in Canada and three other centres. As of June 1981, this constituted coverage of 52% of the Canadian population.

1.2 Other Digital Data Sources
During the past few years there has been a significant increase in the volume of digital spatial data available in Canada, specifically from federal, provincial, and municipal mapping agencies (Tomlinson, 1984). In some cases, digital data is now available for areas within the AMF coverage program. To avoid duplication and reduce overall AMF creation and maintenance costs, Statistics Canada is beginning to look to digital data sources and updating programs of sister mapping agencies. This division of responsibilities compares with the working relationship of U.S. Geological Survey and the U.S. Census in the joint TIGER/U project described by Marx (1983) and Callahan (1984). A number of experiments with other agencies are underway or under discussion. This paper highlights one such experiment conducted with EMR. Other joint projects are listed in Table 1.

Table 1. Agencies Interested in Joint AMF Maintenance Agreements with Statistics Canada

Agency	Agreement	Status
City of Winnipeg	File Maintenance by the City of Winnipeg. Plotting by Statistics Canada.	In production.
Corporation of the District of Burnaby	Burnaby is building a network file with AMF compatible node identifiers. Detailed negotiation is beginning.	Burnaby has submitted a proposal for discussion.
Province of Ontario, Ministry of Natural Resources	Ontario wants to evaluate AMF as a source of attributes and network structure for their topographic data base.	Joint pilot project is underway for the city of Cambridge.
Four Regional Police Forces in Ontario	Coding by clients. Digitizing and processing by Statistics Canada.	AMF creation is progressing well.
Metropolitan Toronto	Metro is maintaining a link to AMF for their planning network file.	A meeting is planned.
C.R.A.R. Inc., Quebec	Network Files for 700 municipalities have been created by C.R.A.R. Incorporated.	Preliminary negotiation is underway.

1.3 The Digital Mapping Programme of Energy, Mines and Resources Canada

The Surveys and Mapping Branch of Energy Mines and Resources Canada has been involved in the automation of the processes of mapping and charting for over twenty-five years. Early activity involved the development of a system for the manual digitizing of 1:50,000 scale map manuscripts and methods for storage, retrieval and automated reproduction of this cartographic data. Parallel to this development, research and development was undertaken into the digitizing of information directly from aerial photography using a hybrid system of photogrammetric instrumentation and in-house developed video-digitizer and software. In both cases it was evident that the collection and manipulation of a complex data set required use of interactive graphics techniques.

Subsequent developments, starting in late 1976 were concentrated in the photogrammetric digitizing approach to data collection with the reproduction capability retained from the manual digitizing developments. This development made use of the interactive graphics capability of Intergraph systems and by late 1978 staff training and engineering development was being undertaken on a production level system.

The production system, which continues to evolve (Gibbons, 1982 and 1984), is predicated on the concept of a digital data file which contains the measured ground positions of terrain features rather than the more usual approach of digitized map coordinates. While this concept requires some additional effort in order to produce a cartographic map, it possesses numerous advantages because it does not suffer the usual distortions, omissions and generalizations of a cartographic product.

Since the beginning of production in the first quarter of 1979, the system has been used to acquire the digital topographic data and produce 1:50,000 NTS map products over some 180,000 km^2. The majority of the effort has been in the more heavily populated area of the province of Ontario following the arc of Lake Ontario from Ottawa - Cornwall to Sarnia.

Based on the results of competitive bids from the surveying and mapping industry, this process of acquiring a digital topographic data base and producing a cartographic map product from this source is about 25% more expensive than producing a cartographic map product by traditional analogue and graphic methods. There is, however, an additional product in the topographic data base acquired.

The utility of the digital topographic data as acquired by the Branch has often been demonstrated by the production of alternate mapping forms. This has included cost effective production of cartographic products for map scales from

1:20,000 to 1:250,000, aeronautical chart bases at 1:250,000 and 1:500,000 and so on. The greater benefit for the approach is, however, expected to be in the non-graphic applications whether as input to mathematical models of the environmental scientist or as a structure to which data sets of spatially distributed themes can be referenced. These non-graphic applications have been slower to develop for several reasons. A significant barrier is the difficulty of transferring digital information without the benefit of established standards for information models, data definitions, data formats, etc. Also the nature of the communication between the producer and the user is changed when exchanging digital data. This is not immediately recognized and imposes certain delays which may frustrate the exploitation of digital data.

There was an identified need to improve the communication with users of traditional map products in respect to digital data. Hence, Surveys and Mapping Branch established a small project team to review with users potential applications for digital data and the needs of these users in relationship to the plans and capabilities of the Branch. The mandate of the project team also included the demonstration of the utility of the topographic data by undertaking small projects with direct assistance to potential users.

Several federal agencies and departments were contacted in the review. However, for various reasons, including the relatively modest amount of data available, only a few demonstration projects were undertaken. A pilot test with Statistics Canada (STC) was selected because of this agency's substantial experience with handling digital graphic files. Furthermore, STC had defined requirements which would provide not only a test of the utility of the data but also an illustration of the characteristics of the data which would be required to meet non-graphic applications.

1.4 The Pre-Pilot Test

Early tests conducted in the summer of 1982 using EMR topographic data for the City of Belleville, Ontario indicated that a number of enhancements were required before application to AMF creation, specifically:

- overshoots and undershoots at intersections of up to 5 meters;
- absence of nodal points at all intersections of 2 or more features; and
- dispersion of logical features (e.g. a street) into several smaller distinct segments not logically connected in any way.

In December of 1982, EMR agreed to create an enhanced file to meet these generic needs for network data as expressed by Statistics Canada. At the same time, Statistics Canada selected the City of Lethbridge for a production pilot trial, since the population of Lethbridge had surpassed 50,000 in the 1981 Census, and an AMF was needed for the

1986 Census. In February 1983, EMR agreed to complete the sheet in question and then to supply Statistics Canada with digital data to agreed upon specifications. Digital files for Lethbridge were received in the summer of 1984, after discussions and software development at both agencies.

1.5 Objectives of the Pilot Test

With the Lethbridge pilot test, Statistics Canada hoped to seriously evaluate the potential benefits and costs, and to identify problems in working together with EMR on digital data projects. The team also hoped to assess potential solutions to these problems and explore them fully with a wide audience before making any permanent changes in procedures.

Statistics Canada recognized a number of potential long term benefits in looking to EMR as a digital data source:

- improved accuracy in existing files;
- improved standardization and compatibility;
- ability to share software development;
- potential unit cost reductions in the long run for new areas; and
- potential compatibile source for elements not currently captured by Statistics Canada such as buildings.

EMR's objectives in conducting the Lethbridge pilot test were as follows:

- demonstrate the utility of its current digital topographic data;
- invite suggestions on improvements to the content and structure of the database;
- assess Statistics Canada as a source of cartographic data; and
- develop an appreciation of the needs and approaches to digital data interchange.

2. METHODS

A methodology was developed which involved development and processing at both departments. The manipulation of the raw topographic data at EMR is described in Section 2.1. The labelling of features and addition of attributes at Statistics Canada is described in Sections 2.2 to 2.4.

2.1 Processing at EMR

The acquisition of the digital topographic data at EMR utilizes techniques and procedures designed to facilitate data collection and subsequent cartographic map production. In consideration of the photogrammetric environment, the data acquisition is mostly unrestricted by factors such as feature sequence and direction. However arbritrary directions or sequences are not readily acceptable as input to the AMF structured file. The critical criteria established for topographic data were the need for completeness, accuracy and a general purpose topographic feature coding scheme which describes the 'real world' characteristics of a feature.

The three major problems associated with the transformation of the basic topographic data to a form more amenable to the requirements of the AMF are: the arbitrary sequences and segmentation of features, absence of coordinates within a string to define intersecting nodes, and overshoots or undershoots. It was also necessary to consider a major difference in the nature of a network approach to data collection from the approach required to ensure an accurate portrayal of features in a multi colour map. This latter point is manifest in the number of vertices used to describe linear features and brings with it the need for appropriate generalization of linear features.

First, the requirement to reorder and reformat data for a variety of applications had already been recognized at EMR. A process was developed to reorder, logically connect and generalize linear features. The process is table-driven with respect to the feature coding scheme and permits the application of distinct criteria for different feature types. The control parameters which can be specified for this process are (i) a truncation value, which defines the minimum permissible distance between successive points in a string, (ii) a curving tolerance which defines the permissible rate of curvature of a string, and (iii) a closure tolerance which defines the maximum distance between successive elements in a chain for the chain to be considered continuous.

The second requirement of intersecting nodes involves three operations: determining the actual intersection of features, rebuilding the line string or coordinate chain to include the determined intersection coordinates and finally, flagging of the intersections or nodes. The process for building the node file for Statistics Canada was coded to operate in a PDP 11/70 environment and has a flow as follows:

- sort all selected features segments by the min-max of the end point coordinates;
- string the segments together into a complete feature after application of the truncation, curving tolerance, and closure parameters;
- starting from the feature with a minimum x-y, determine intersections with successive features, including the cases for undershoot, and build an index of intersections;
- using the index of intersections, process each curving feature to incude new intersection points and handle the cases for overshoots and multiple intersections;
- reprocess the index of intersections as modified for multiple intersections into a single list.

The resulting file should now consist of a number of coded features with intersection points appropriately inserted in each chain and the nodes separately identified as point features. The separation of linear features from the socalled "node features" was necessary to retain a graphics capability for plotting the file at EMR, although a separate data structure combining the coordinate chain and the node flags would be a preferred end product.

The full processing of the pilot test includes extraction from the topographic data base, according to the geographical extent and the feature selection made. This extraction process was only performed once and the results retained as an interim file. The time required to process this interim file for noding was approximately 25 minutes. The transfer to magnetic tape interchange format is very straightforward. The package transferred to Statistics Canada included the formatted tape, and summaries of the features transformed, number of vertices and the like.

2.2 Overview of Processing at Statistics Canada

The primary function of the Interactive AMF Creation System (IACS) is to build an AMF from an EMR topographic feature file and various source documents.

An EMR topographic feature file describes a network of features for a designated area. Each feature is defined by a series of points linked together. One element of descriptive data is also supplied for each feature, a 4-digit feature code that identifies its feature-type according to an EMR classification scheme.

Source documents consist of assorted street lists and maps obtained from the municipality and various other agencies. These documents contain the descriptive data, or attributes, associated with each feature such as NAME, DIRECTION and ADDRESS-RANGES. Since this data is not always available on one map, some attributes must often be obtained from other sources. With Lethbridge for example, address-ranges were not available but postal codes were. Therefore, an extra manual processing step was required to translate postal codes into address ranges. These source documents are also used to supplement the EMR feature network by digitizing additional features such as the municipal boundary. Streets may also be digitized if the source document contains more recent data than the EMR file. The output AMF describes the street network and other physical features for a designated area. Each feature is defined by a series of nodes linked together and is described by such attributes as NAME, FEATURE-TYPE, SUB-FEATURE TYPE, DIRECTION, etc.

The IACS must, therefore, produce an AMF by merging the feature-network received from EMR, with the descriptive data obtained from source documents.

2.3 Detailed Processing at Statistics Canada

The IACS consists of 7 major processing steps. Each process is described below.

A. LOAD - NETWORK - FILE

- the feature network is loaded by a special interface program into a model on an interactive mapping system (AUTOMAP).
- in the process, EMR feature- types are translated into STC internal feature-types and EMR ground co-ordinates are converted to UTM.
- the model is validated by an online spot check to ensure that there are no "overlaps" or "undershoots" and that a node exists in both features where they intersect.
- network plots are produced at BASE-MAP scale for final validation and for transcription of feature-names and other attributes.
- at the end of this step, a model exists containing a description of the feature network without any attribute data.

B. ANNOTATE - NETWORK - PLOTS (Map compilation)

- all feature attributes, including feature name, street type, start point, direction and feature type are coded on the NETWORK-PLOTS from the most recent municipal source documents in preparation for input to the feature network model.
- address ranges are excluded because they are node-attributes and including them interactively would make inefficient use of the cartographic workstation.
- missing features, such as the municipal boundary and new streets, are also drawn onto the network plots for input to the feature network model through digitizing.
- features such as small ponds and trails, which were contained on the EMR file but not required for the AMF, are highlighted in yellow.
- at the end of this step, the network plots are annotated with all feature attributes in a formatsuitable for input to the network model.

C. CREATE COMPOSITE NETWORK MODEL

- a composite feature is created for each desired feature in the model, containing the feature attributes and all of its nodes.
- customized macros were developed to prompt operators for all required information and to provide instructions on procedures.
- missing features are digitized to complete the model.
- some features must also be split if the name changes.
- at the end of this step, the COMPOSITE NETWORK MODEL contains nodes and attributes for all required features.

D. PRODUCE MINI - AMF

- the COMPOSITE NETWORK MODEL is transformed into AMF format and transported from the HP 1000 mini-computer to the the AMDAHL V8 mainframe for further processing.
- the transformation includes creating an AMF FEATURE for each COMOSITE FEATURE and translating internal AUTOMAP feature types into AMF feature types.
- some generated fields such as feature code and sequence number are also calculated.
- some validation is also done and errors are corrected using the interactive graphic editing system.
- at the end of this step, an AMF exists, although some required fields are missing.

E. COMPLETE MINI - AMF

Three major operations are performed on the MINI AMF in this phase by batch processing.

1) LABEL NODES
   - node and section numbers are calculated for each detail record based on x and y coordinate UTM values
   - section numbers are assigned so that eachsection corresponds to an NTS map sheetof scale 1:5000.
   - node numbers are then assigned sequentially within each section, in ascending order based on their x - y location
2) RESEQUENCE FEATURES IN ALPHABETICAL ORDER
   - each AMF feature has a unique 6 digit feature code associated with it.
   - the features are sorted by feature - name and feature codes are reassigned sequentially in ascending order i.e. alphabetical order
3) INSERTION OF CROSS-FEATURE IDENTIFICATION
   - each detail record is linked with other detail records that have the same node and section numbers.
   - these represent intersections and the feature name, feature code and feature type of the intersecting feature are recorded in the cross reference fields of each detail record.

At the end of this step, the AMF is complete,except for address ranges and block-face centroid values.

F. INSERT ADDRESSES

Address ranges must be added to the detail records of all addressable features that represent BEGIN, END or INTERSECTION nodes. This process involves several steps as follows:

- if a base-map with address ranges does not exist, then one is compiled from available data.
- a special format printout of the AMF is produced for coding address values.

- coded addresses are then keyed to create a file of AMF update transactions.
- these transactions are processed by standard AMF update software which places address ranges in the AMF and calculates block-face centroid values.

At the end of this step, the AMF is complete except for final validation.

G. VALIDATE AMF

Several steps are taken to validate and if necessary correct the AMF:

- plots are produced for visual verification.
- special programs are run that perform various quality checks.
- final verification is done by a special program that creates ADD-TRANSACTIONS for all AMF features. These transactions are then run through the standard AMF update procedure that contains all the validation rules for the AMF.
- any errors identified are corrected using standard AMF update procedures.

At the end of this step the AMF is complete and available for production use.

2.4 Development at Statistics Canada

One of the attractive features of the IACS was that most of the major components already existed in some form and needed only minor modification to implement IACS function.

The interactive graphic editing system at Statistics Canada, AUTOMAP, had already been used to edit and plot AMFs in both the Computer Assisted Collection Mapping project (Yan, 1982) and the Interactive AMF Update System (IAUS).

For Phase A, loading the feature-network, an existing program developed for the Belleville pre-pilot test was used. The EMR to AUTOMAP program (EA001) needed only minor adjustments to the feature table. Improved run statistics and more informative diagnostic messages were also added.

Phase B, annotating the network plots, is a clerical procedure performed by the Geocoding Unit who have a great deal of experience with similar tasks. Consequently little additional training was required and the only development needed was the documentation of a few procedures.

The third phase, creating the composite network model, was implemented using the facilitities available with the AUTOMAP system. The AMF file structure had already been implemented under AUTOMAP in the IAUS project. AUTOMAP macros were developed to prompt operators for required data from the network plot and to instruct them on how to proceed. Procedures were also documented to augment the online facilities.

Phase D, producing the Mini-AMF, was implemented using the Mini to AMF program (MA001), developed in 1981 for the updating of the St. Catharines AMF as part of the IAUS.

However the MA001 program had to be modified extensively to process an entire AMF and to generate missing fields which were normally present in update mode.

Completing the MINI-AMF in PHASE E was implemented as a mix of old and new. The generation of Node and Section numbers, based on X Y co-ordinates, was an experiment with possible far-reaching impact on the future of the AMF file structure. In the traditional AMF creation method, node numbers are arbitrarily assigned and the AMF is split into sections irrespective of any regular geographic pattern. The section and node numbering algorithm implemented in the IACS ensures that each section corresponds to a fixed 1:5,000 scale map sheet, and that each node number could be calculated based on its x-ycoordinates. If successful, this experiment would demonstrate that it is feasible to create and maintain AMFs by sections representing stable geographic areas as opposed to areas affected by dynamic political boundaries. This method was implemented as a new program and was based on work at the Centre de Recherche en Amenagement Regionale (CRAR) at the University of Sherbrooke.

Resequencing and chaining were implemented using existing AMF programs with virtually no new development.

Phase F, inserting addresses, was implemented based on existing AMF update procedures. A few small utilities were developed to produce the special format print for coding addresses and for capturing and processing the address values. The captured data are formatted into standard AMF update transactions which are processed by existing AMF updates program. Procedures for coding the address ranges were also developed.

Validating and correcting the AMF in Phase G, is done using existing programs. Virtually no development was required because the existing sub-system contains sufficient quality checks. This process was however augmented by various acceptance tests at the end of each phase.
In summary, the IACS is functionally similar in many respects to the traditional AMF creation system with the primary difference being the source and format of the input data. The major difference in implementation is the use of interactive graphic hardware and software to replace the traditional manual coding method.

3. RESULTS

3.1 Lethbridge Test Results

To this date, approximately 40% of the Lethbridge data (240 of 600 features) have been processed up to and including final validation of the AMF. The quality of the end product is acceptable except for nodes missing at many intersections. This was caused by a recently detected software bug at EMR and highlights the need for improved quality checking of digital data received from other agencies. Other minor problems have been identified, all of which can be corrected with small software modification.

Results to date demonstrate that the fundamental approach, the functional flow of the process and the underlying algorithms developed are effective and viable in creating a data structure suitable for network applications. Nevertheless, further testing and quality assurance is required.

Table 2 compares the estimated cost of producing the test AMF using traditional methods with actual timings achieved so far. The interactive data exchange method seems to offer savings in terms of cost and elapsed time in AMF creation. Furthermore, bottlenecks have been identified and fine tuning of the software will result in further savings, particularly in the interactive labelling of features.

Table 2. Comparison of Time for Area Master File Creation - Normal Process vs. IACS Data Interchange Process Based on Preliminary Results

Traditional Process Activities	Planned Person Days	Actual Person Days	IACS Data Interchange Process Activities	Planned Person Days	Actual Person Days
Create address document	6.25	11.0	Load network file	2.0	3.0
Base map compilation	6.25	5.0	Annotate network plots	6.25	5.0
Node symbolization	5.5		Create composite network model	15.0	*25.0
Node numbering	5.5		Produce Mini AMF	2.0	2.0
Coding	22.0		Complete Mini AMF	3.0	2.0
Assign central point	.75				
			Insert Addresses	20.0	**20.0
Digitizing	2.75		Validate AMF	5.0	5.0
Analyse Run 01 results	.15			53.25	62.0
Analyse and correct section plot	11.0				
Analyse and correct node plot	5.5				
Analyse and correct centroid plot	2.1				
Finalize AMF	2.5				
	70.25	72.75			

* 25.0 days based on completing 600 features using the rate achieved during the last week of production. Training and development time is not included.

** Estimated cost for address insertion.

3.2 Differences Encountered

As well as demonstrating the feasibility of digital data exchange between EMR and STC, the pilot study also identified several areas of incompatibility that must be addressed in the long term.

Many problems were encountered reconciling EMR and AMF feature classifications. For example, EMR recognizes over 20 road-types, based primarily on relative size, whereas AMF recognizes only 5. On the other hand, since EMR has established very few classes for hydrographic features, STC could not be precise enough in requesting the specific features types required. Also, the criteria for classification were not defined sufficiently clearly by either agency, thus further complicating the task.

Feature representation is another area of difference. Whereas EMR represents a divided road or highway as 2 separate features, AMF stores only 1. EMR represents a road with a bridge on it as a continuous feature. AMF, on the other hand, stops the road at the bridge and continues on the other side, effectively splitting the feature. The AMF also maintains some features not maintained by EMR such as municipal boundaries. These features must be added to the AMF model as a separate step.

Requirements for currency of the data also vary for each agency. The AMF is tied strongly to the Census and therefore must be as up to date as possible every 5 years for Census day. New subdivisions identified since the EMR data was captured must be included in the AMF to ensure it is as complete as possible for Census uses.

Finally, the basic unit of work is different. EMR bases their work on standard NTS map sheets, while AMF uses a politically defined municipality as a unit of work.

3.3 <u>Guidelines for Future Work</u>

In the long term, a number of points seem critical to facilitate this type of joint base file maintenance program

(1) Move towards a compatible classification scheme for features at both agencies.

(2) Agree on a common coordinate encoding scheme for feature e.g. boulevard streets, creeks, railway yards, shoreline, bridges.

(3) Develop a standardized transfer format which includes not only topography (coordinates and feature codes) but also topology (nodes, links and areas) and attributes (feature names for example).

(4) Develop a common set of conventions regarding treatment of networks, including nodes, behaviour at intersections, what can cross without a junction e.g. powerline and road, bridges and roads.

(5) Level of generalization should be agreed upon (e.g. in railway yards, how many lines should be coded?)

(6) The final division of responsibility needs to be agreed upon, but it appears that EMR should be the primary agent for distributing basic topography, with other agencies such as Statistics Canada responsible

for recording attribute data and some layers including geo-statistical boundaries.

(7) Planning and scheduling of the complete process beginning with aerial photography should involve significant consultation to optimize currency of critical data elements.

(8) Ongoing consultation by all participating agencies including Statistics Canada is required on a number of aspects including:
- delineation and classification
- transfer format
- standards
- updating requirements

3.4 Action Taken In Implementing Guidelines

There is already progress on a few of the items listed above.

(1) Statistics Canada's AMF staff are planning to review the AMF feature classification against the EMR standard.

(2) A federal-provincial committee has been formed to examine a standard digital transfer format with consideration of topology, attributes and topography based on extending the proposed standard of the Canadian Council on Surveying and Mapping (1982).

3.5 Open Questions

A number of questions remain that must be seriously examined. Recognizing their importance, these questions are provided here to give an indication of future directions.

- Do users always want data by map sheet?
- Should data be available by administrative boundary as well e.g. county or township?
- What should the links be across map sheets?
- What should the links be across administrative boundaries?
- How should digital data be packaged geographically?
- How should digital data be packaged in terms of level of feature?
- How should digital data be packaged in terms of topology?
- Who are the primary users of this digital data?
- How much are they willing to pay?
- How should updating of shared data be handled?

4. CONCLUSIONS

Early results from a pilot test indicate probable benefits to both agencies from continuing joint efforts in digital capture, application and transfer of street network data. A list of suggested procedures and important issues for inter-agency base file maintenance projects has been provided for consideration.

ACKNOWLEDGEMENTS

The authors wish to thank Catherine Gourley, Suzanne Enright, Rejean Fontaine, Andre Landreville and Mark Doherty for assistance with the programming and Carl Sadler and Mila Cardenas for production. The support of Ross Bradley, Andre Boisvenue and J.M. Zarzycki in this interdepartmental venture is also acknowledged.

REFERENCES

Boisvenue, A. and Parenteau, R., 1982, "The Geocoding System in Canada and the Area Master File", Papers from the Annual URISA Conference, pp 226-232.

Broome, F.R., 1984, "TIGER Preliminary Design and Structure Overview: The Core of the Geographic Support System for 1990", presented at 1984 Annual Meeting of the Association of American Geographers , Washington, D.C.

Callahan, George M. and Broome, Frederick R., 1984, "The Joint Development of a National 1:100,000-Scale Digital Cartographic Data Base", Proceedings of the American Congress on Surveying and Mapping, Washington, D.C.

Canadian Council on Surveying and Mapping, 1982, National Standards for the Exchange of Digital Topographic Data, Energy Mines and Resources Canada.

Gibbons, J.G. 1982, "Digital Techniques for 1:50000 Scale Mapping and Revision", Photogrammetric Record, 10(60) pp. 645-652, October 1982.

Gibbons, J.G. 1984, "An Operational System for Digital Topographic Data Acquisition and Digital Cartography", 15th ISPRS Congress, Rio de Janiero.

Marx, Robert W. "Automating Census Geography for 1990", American Demographics, June 983.

Statistics Canada, 1972, GRDSR: Facts by Small Areas.

Tomlinson Associates, 1984, Investigation of Digital Cartographic Status and Developments in Canada, 5 Volumes, DSS Contract No. 03SQ.23246-4-5508 for Energy, Mines and Resources Canada.

Yan, J.Z., and Bradley, D.R., 1982, "Computer Assisted Cartography for Census Collection: Canadian Achievements and Challenges", Proceedings of the Sixth International Symposium on Automated Cartography, Volume 2, pp 135-146, Ottawa.