THE DEVELOPMENT OF A DATA MODEL AND NATIONAL STANDARDS FOR THE EXCHANGE OF DIGITAL TOPOGRAPHIC DATA

Dr. M.M. Allam

Senior Science Advisor
Surveys and Mapping Branch
Energy Mines and Resources Canada
615 Booth Street
Ottawa, Ont., Canada K1A 0E9

ABSTRACT

Digital mapping systems have different origins, application scope and design philosophy. These systems were developed to supplement data acquisition systems , to serve as a data base system, or as an interactive graphic system with application software for cartographic purposes.

Under the auspices of the Canadian Council on Surveying and Mapping (CCSM), national standards for topgrapgic data interchange were developed in Canada. These were required to facilitate the interchange between different systems for various operational and economic reasons, and for the timely dissemination of digital data among federal, provincial and private organizations.

This paper discusses the activities leading to the development of the national Canadian standards for: the classification and coding of digital topographic data, the quality evaluation of the digital topographic data, the digital topographic data model and the EDP File Format for the transfer and interchange of the digital data.

INTRODUCTION

During the past two decades mapping organizations have witnessed a tremendous growth in digital mapping and automated cartography systems. Consequently, the mapping community has seen a proliferation of various approaches to the acquisition, coding and storing of digital topographic and cartographic data. As a result there are as many data bases, with specific file structures as there are vendors.

As might be expected, each vendor of a computer aided mapping system has a view point on what should be standardized and how. And, as might be expected these views are coloured by proprietary considerations. This explains why data interchange between different systems is virtually impossible without a transitory file format, and software for the encoding/ decoding of the data.

Recognizing the fact that the continuing prolifertaion of differing approaches to the handling of digital topographic data could hinder the orderly development of the national digital topographic data base, the Canadian surveying and mapping community became concerned with the lack of standardization in the handling, storing, retrieving, transmitting and merging of digital spatial data. In May 1978, this concern led the Government Task Force on National Surveying and Mapping to recommend that the Surveys and Mapping Branch develop national standards for the interchange of digital topographic data.

In October 1978, under the auspices of the Canadian Council on Surveying and Mapping (CCSM), three technical committees were formed for the development of standards for: the classification and coding of digital topographic data, the quality evaluation of the data, and a file format for data interchange.

The first draft of the standards was published in April 1982[CCSM, 1982] and was distributed to more than 700 organizations and individuals in Canada and abroad. The recipients were asked to comment on the standards and these served as the basis for publishing the second draft in October 1984.

The developed standards for the classification/ coding of the data, and the standards for quality evaluation were approved by CCSM in November 1984, and were generally accepted by the Canadian mapping community.

The developed file format was not so fortunate. For its development, the Committee used the Computer Compatible Tape CCT Superstructure concept, as defined by the LGSOWG (Landsat Ground Station Operators Group). The CCT superstructure is a self defining format, and as such contains some very rigid specifications. It contains significant detail about the specific files of data to permit the reading of the information without prior knowledge of its source.

This flexibility was accompanied by a certain degree of complexity, which added additional constraints towards the acceptance of the CCSM file format. In addition to the difficulties encountered with the CCT superstructure concept, the developed file format lacked the existence of an efficient mechanism for the transfer of the topological information and the spatial relationships between the topographic features.

In October 1984, under the auspices of the CCSM, a meeting was held to discuss a proposal by the Surveys and Mapping Branch of the Ontario Ministry of Natural Resources for the development of a revised format. This meeting resulted in the formation of a technical subcommittee consisting of representatives from the federal and provincial Surveys and Mapping Branches, the Canada Centre for Remote Sensing and Statistics Canada. The subcommittee was tasked with discussing the merits of the CCSM file format vis-a-vis the Ontario proposal. In November 1984 the members of the subcommittee recommended that a digital topographic data model be developed and that the CCSM File Format be revised or redesigned to include topology and attribute information. In addition, it was recommended that the redeveloped file format be simple to understand and use by the digital mapping community.

Considering these factors a working group of the technical subcommittee was formed in 1985 for the development and/or revision of the digital topographic information model (DTIM) and the CCSM file format. The DTIM report was published in November 1985 and the new CCSM File Format was developed in May 1986.

This paper describes the data classification/ coding standards and the digital topographic data quality standards, and addresses the major components of the developed digital topographic data model and the new CCSM File Format for the interchange of digital topographic information.

STANDARDS FOR THE CLASSIFICATION/ CODING OF FEATURES

Topographic Feature Classification

For the classification of topographic features, a tree structure approach consisting of four levels was adopted as follows: I) Class , II) Category, III) Feature and IV) Attribute. In level I the following major classes were identified: Designated Area, Building, Structure, Roadways and Railway, Utility, Delimeter, Hydrography, Hypsography and Land Cover. Level II represents the category of information, which is a break-down of the Class level. For example under Building we find categories such as, commercial, governmental, residential, etc.. Topographic features were included on level III and their non-graphic essential attributes in level IV.

Coding of Topographic Features

For the coding, alphabets were used for the Class and Category levels, while five digit integer codes were used for the features. The essential attributes in level IV, which forms an integral part of the feature were allowed a three digit integer code. Detailed information on the classification and coding is contained in Part I, of Volume I of the CCSM stndards (CCSM, 1984), and a complete listing of all the topographic feature codes is included in Volume II of the standards. In addition, Volume II includes a dictionary of the terms which describe clearly and unambigously, all the classified and coded features.

STANDARDS FOR QUALITY EVALUATION OF DIGITAL TOPOGRAPHIC DATA

These standards mainly included a set of rules by which producers of digital topographic data will be able to calculate realistic estimates of the accuracy of the data contained in their files. To achieve this, an empirical formula for accuracy calculation was developed. Also, a practical method for evaluating the accuracy of the digital mapping process was included.

Other issues addressed in these standards were: the definition of all the parameters required for the management, processing, and retrieval of the data. In addition, the data quality standards included a review of the prominent mapping accuracy specification, and a review of digital data acquisition techniques and technologies. A detailed account of these standards was included in Part II, Volume I (CCSM, 1984).

DIGITAL TOPOGRAPHIC DATA MODEL

Definitions and Basic Data Components

In general data corresponds to discrete, recorded facts about phenomena from which we gain information about the world. A data model defines the rules according to which data are structured. In the context of digital topography data may be considered as the passive entities or objects of an activity. In digital topographic mapping, or in the process of creating a national digital data base, the "Topographic Feature" is an object on the earth's surface and it usually has some significance. Examples of features include a building, road, lake, park or contour line. These features are real world entities of interest that have been defined. As stated before features are grouped into Categories and the Categories into Classes.

In geo-based digital data bases, the basic data components that are required to describe the feature are the spatial coordinate (locational) data, the non-graphic attribute or descriptive data, and the topological and spatial relationships among the features.

Spatial data defines the position of the feature or group of features in space using a reference system relative to known locations on the earth's surface. Spatial extent, geographical extent, and absolute position are all terms synonymous with location. Points are represented by a single coordinate set, and lines are represented by two or more connected points. Areas are represented by their relation to the lines which bound them. Other components associated with spatial data include resolution, positional accuracy, and projection (coordinate system specification), and these are normally given as non-graphic attribute data.

Attributes are non-locational and non-topological data attached to a feature, to a category, or to a class of features in the data set. This data consists of discrete feature codes (CCSM feature classification code), and a set of user defined attributes such as: textual descriptors, label identifiers and any other phenomena as required to describe the feature. For example, a feature such as road, may have a set of user defined attributes which may include information on: surface material, number of lanes, jurisdiction, road label, road number, speed limit, status, etc.. In addition each feature must have a unique feature identification number for quick referencing in data file.

Topological data includes information related to topological structure of the data and the spatial relationships among the various features on the earth's surface. Also, it provides information on the neighborhood of a feature, that is physically contiguous, to a given feature.

Topology in the Model

Topology is concerned with nodes, lines and areas and their inter-relationships (connectivity and adjacency). In the model only two primitive feture types were used : point and line. Other feature types that may be derived from these primitives are: node, segment or chain, and polygon or area. Therefore, nodes are of interest as end points of lines and as vertices of areas. Areas are formed by lines and are adjacent to other areas. This is the primitive topological information that may be captured with digital topographic information.

Spatial relationships between the features are simply the primitive linkages by which associations among the features are described. For example, the feature "road" may cross another feature such as "railroad". In this example the spatial relationship between the features is "Cross", and this is frequently found on networks of intersecting lines.

Some relationships between topgraphic features are quite simple such as the concatenation of two linear features of the the same class (e.g. two streams in the Hydrography Class which normally consists of merging/intersecting networks of lines). Other relationships are more complex such as the overlapping of areal features belonging to different classes. For example, the overlapping of "forest" area boundary with the boundary of a "county area". In this example the relationship "overlap" is between the "Land Cover" and "Delimeter" classes.

A potential set of relationships between features include the example set given in table 1 (Goodchild, 1985).

point - point	Is nearest to ..
	Interacts with ..
	Is allocated to ..
Line - Line	Joins ..
	Is upstream of ..
Area - Area	Is adjacent to ..
	Interacts with ..
Point - Line	Is on ..
	Is nearest to ..
Point - Area	Is within ..
	Is allocated to ..
Line - Area	Crosses ..
	Borders ..

Table 1. Example of Spatial Relationships

The above relationships were cited as an example only and they may expand depending on the required complexity between the various features in the same category, or various classes of feaures.

FILE FORMAT

Based on the developed digital topographic information, and after a thorough review of the EDP standards developed in 1982, a new file format was developed by Perceptron Computing Inc., Toronto, Canada, in May 1986.

Basic Requirements of the Transfer Format

In developing the transfer format, the following basic requirements were considered:

- machine and language independence, to allow for the exchange between different computer system;
- simple to use and understand;
- allowing for self defining, e.g. by having a standard internal format wherein structures are defined at the start of the data sets and/or data sub-sets which decribe the remainder of the data;
- modularity and expandability, to handle new types of data;
- flexibility, by allowing for options in data components, i.e. giving the user options for the transfer of attributes and/or topology or no-attributes and/or no-topology, i.e. the user may have topology and/or attributes or none.

Structure of the Data Exchange Format

As shown in Figure 1, the Data set consists of:

a) a Volume Directory File (VDF)
b) 1 to N Topographic Data File(s) (TDF)

The VDF and each TDF are followed by an end-of-file (EOF) mark, and the last TDF on the logical volume will be followed by an (EOF) mark. Also, allowance is made for files to span physical volumes.

The VDF consists of the Volume Descriptor Record (VDR), which contains EDP general information for the identification and management of the physical volume.

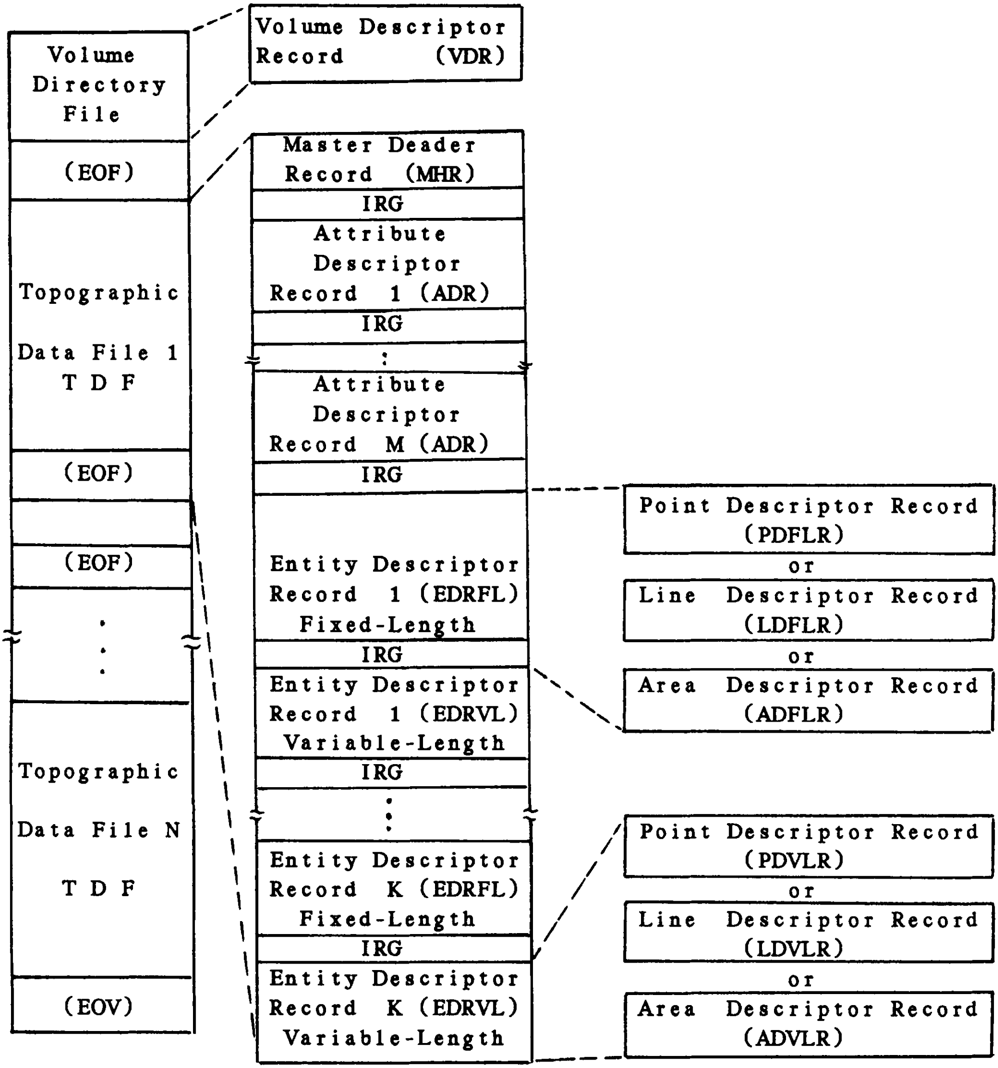

Figure 1. File Format Data Interchange

Each TDF consists of:

a) Master Header Record(s) MHR
b) Attribute Descriptor Record(s) ADR
c) Entity Descriptor Record-Fixed Length, EDRFL
d) Entity Descriptor Record-Variable Length, EDRVL

The MHR includes information of file name, date of data collection, feature type in the TDF (point, line, area), type of topologically-related data (point, line, point-line, line-area or point-line-area), range of data in the TDF and information specifying the projection, the coordinate systems and units.

Each ADR contains a group of attributes consisting of an attribute name, attribute type identifier and the length of character string for each attribute.

The EDRFL or EDRVL records vary depending on whether the record is for a point, line or an area.

The Point Descriptor Fixed Length Record (PDFLR) contains point I.D. number, spatial coordinates and the number of attached lines and the attribute values as per order in the ADR.

The Point Descriptor Variable Length Record (PDVLR) consists of line(s) identification number(s) attached to the point.

The Line Descriptor Fixed-Length Record (LDFLR) consists of the line I. D. number, start and end point I.D. numbers, left and right area I.D.(s), the number of coordinates in the line and the attribute values as per their order in the ADR.

The Line-Descriptor Variable Length Record (LDVLR) contains spatial coordinates in the lines.

The Area Descriptor Fixed Length Record (ADFLR) includes the area I.D., coordinates of a centroid in polygon, number of lines in boundary and attribute values as per their order in the ADR.

The Area Descriptor Variable Length Record ADVLR contains the I.D. numbers of all boundary lines.

CONCLUSION

The history of data exchange standards is one of evolving solutions. These solutuions came early in the form of de facto standards to answer the hardware and software compatability problems brought about by the proliferation of digital mapping and computer assisted cartography.

The developed National Stardards will serve as the basis for the exchange of digital topographic data between Federal, Provincial and Private surveying and mapping agencies in Canada. The introduction of these standards will reduce costs through pooling the topographic information collected by the Canadian mapping organizations into an entity in the national data base, thus minimizing duplication of work, and also it will help minimize the expense of exchanging data between mapping agencies. However, it is not foreseen that survey organizations will necessarily operate and/or store data in this exchange format. The main purpose is to provide a national format and to establish guidelines for the exchange of digital topographic data. It is aslo hoped that with the adoption of the national standards, the days of end-users depending on one vendor for their system needs are over.

The developed file format may be the basis for reaching an agreement among the different vendors toward the adoption of standardized file format. This format can be directly used for the migration of digital cartographic data from one system to another, without the need for a translator software.

We have seen vendors responding favourably to the standardization of data communications among different computer systems, and standardization of computer graphics systems and commands, and there is no reasons for the vendors not to consider the archival of data files in their system in a "Standardized" File Format.

REFERENCES

1. Allam M. M., 1983. "National Standards for the Exchange of Digital Topographic Data", FIG XVII International Congress, Sofia,Bulgaria, June 19-28, 1983, Vol. 5, pp. 512.4/1 - 512.4/10.

2. Canadian Council on Surveying and Mapping, 1982. "National Standards for the Exchange of Digital Topographic Data. Draft Reports: I - Standards for the Classification of Topographic Features; II - Standards for the Quality Evaluation of Digital Topographic Data; III - EDP Standards Applied to Digital Topographic Data". Ottawa, Surveys and Mapping Branch, EMR, Canada.

3. Canadian Council on Surveying and Mapping, 1984. "National Standards for the Exchange of Digital Topographic Data. Second Draft Report: I - Data Classification, Quality Evaluation and EDP File Format; II - Topographic Codes and Dictionary of Topographic Features". Ottawa, Surveys and Mapping Branch, EMR, Canada.

4. Canadian Council on Surveying and Mapping, 1985. "Proposed Standards for a Digital Topographic Information Model - W. G. I, Technical CCSM Sub-Committe, Draft No. 2, November 1985.

5. Goodchild, Michael F., 1985. "Geographic Information Systems in Undergraduate Geography: A Contemporary Dilemma", paper presented at CAG Annual Meeting, Trois-Rivieres, May 27-31, 1985.

6. Goodenough, David G., et al. 1979. "Standard Format for the Transfer of Geocoded Information in Spatial Data Polygon Files", CCRS Research Report 79-3, Ottawa, EMR Canada, Canada Centre for Remote Sensing.

7. Moellering, Harold, editor, 1985. "Digital Cartographic Data Standards: An Interim Proposed Standard". Issues in Digital Cartographic Data Standards, Report No. 6, Columbus, Ohio, National Committee for Digital Cartographic Data Standards, 164 pp.

8. Perceptron Computer Inc., 1986. EMR, Topographical Standards, EDP File Format, Draft of an unpublished manuscript, March 1986

9. Stardards Association of Australia, 1981. "Interchange of Feature Coded Digital Mapping Data", Australian Standard 2482-1981, North Sydney, N.S.W.: Standards Association of Australia, 24 pp.

10. Tuori, Martin and Moon, George, 1984. "A Topographic Map Conceptual Data Model", Proceedings of the International Symposium on Spatial Data Handling, Zurich, Geographisches Institut, Universitat Zurich-Irchel, pp. 28-37.

11. U.S. Geological Survey, 1983. "USGS Digital Cartographic Data Standards", Circulars 895-D, Digital Line Graph and 895-G DLG Attribute Coding Standards, Reston, Virginia, U.S. Geological Survey.