ISSUES IN DIGITAL CARTOGRAPHIC DATA STANDARDS

Report #3

Digital Cartographic Data Standards: Defining the Issues

Completion of the First Year of Work

Harold Moellering, Editor

July, 1983

National Committee for Digital Cartographic Data Standards
Numerical Cartography Laboratory
158 Derby Hall
Ohio State University
Columbus, Ohio
U.S.A. 43210

The Committee operates under the auspices of the American Congress on Surveying and Mapping. This work of the Committee is being funded in part by U.S. Geological Survey Grant #14-08-0001-G-807. Production assistance has been provided by the Ohio State University Research Foundation (RF Project #715367).

PREFACE

This report is the third in the series which discusses the work of the National Committee for Digital Cartographic Data Standards. It contains five papers which define the issues involved in order to establish digital cartographic data standards for the United States. The first paper by Moellering discusses and provides an introduction and background to the issues while the remaining papers by Nyerges and White, Chrisman, Schmidt, and Edson provide detailed discussions of the issues from the Working Groups. These papers are expanded discussions of oral presentations originally given at the 1983 Spring meetings of the American Congress on Surveying and Mapping in Washington, D.C.

This report represents the work of the Committee for the first year of operation, that of defining the issues. We now invite public comment on the issues as presented and discussed herein. Please note that there are five sheets in the back of this report where one can provide comments and opinions for the consideration of the committee. Please return all comments in writing to the committee at the address on the title page. Please note that only written comments can be processed by the Committee due to limited staff and resources.

Harold Moellering
Series Editor

TABLE OF CONTENTS

ISSUES FACING THE DEVELOPMENT OF DIGITAL CARTOGRAPHIC DATA DATA STANDARDS: BACKGROUND AND INTRODUCTION

BY

Prof. Harold Moellering
Numerical Cartography Laboratory
158 Derby Hall
Ohio State University
Columbus, Ohio
U.S.A. 43210
Tel: (614) 422-0468

ABSTRACT

The National Committee for Digital Cartographic Data Standards has been fully organized and in operation for a year. It is progressing towards the goal of developing digital cartographic data standards in cooperation with the profession. The first cycle of defining the issues has now been completed and is discussed in detail in the following papers. Work is now continuing on examining the alternatives which will be discussed in future reports.

I. INTRODUCTION

In the last decade the rise of computer-assisted processing in cartography and the use of cartographic databases have grown in a very dramatic way. In the early years agencies in the governmental sector and organizations in the private sector used these approaches on a largely experimental basis. Today one sees the same organizations and agencies using numerical processing and data bases on an ongoing day to day basis. This growth and development in the field of cartography is nothing short of remarkable. However this growth has been uneven and to a large extent not well coordinated in the civilian sector of cartography. It was recognized several years ago that if continued growth in the use of numerical methods and databases was to be sustained over a longer time period, that it would be necessary to develop standards for digital cartographic databases.

Each year millions and millions of dollars are being spent to reorganize, reformat, process, verify and check digital cartographic databases that one agency or organization obtains from another. In most cases this is a very time consuming process. If a comprehensive set of national digital cartographic data standards can be developed, then the entire field of cartography will benefit from this effort and at the same time save large amounts of time and money. To this end the National Committee for Digital Cartographic Data Standards has been established under the auspices of the American Congress on Surveying and Mapping, with the support of the U.S. Geological Survey.

This paper discusses the work of the Committee in its first year of operation, that of defining the issues involved. At the outset the paper explores the evolutionary process that led to the need for digital cartographic data standards. It will also discuss the current work of the committee and introduce the four following papers that discuss the specific issues in detail.

II. RECOGNITION OF A NEED FOR COORDINATION AND STANDARDS

Explicit recognition of the need for coordination of effort in digital cartography and the need for standards has existed for at least a decade. In a 1973 report the Federal Mapping Task Force recommended a greater coordination of cartographic activities in the Federal sector of the profession and also recommended the establishment of a national digital cartographic database (Federal Mapping Task Force, 1973). More recently the National Research Council, an arm of the National Academy of Sciences, in 1980 in their recommendation for a multipurpose cadastre stated that:

> We recommend that technical studies continue to be sponsored by the federal government to identify consistent land information and display standards for use among and within federal agencies and between federal and state governments. These studies should rely on the authority of state governments to adopt the standards and organize the data collection in cooperation with the federal government to ensure compatibility on a national basis, delegating these functions to local governments where appropriate.

In 1981 a different National Research Council panel reviewed the original Federal Mapping Task Force recommendation and subsequently stated that:

> We recommend that the mapping, charting, geodesy, surveying and cadastral agencies of the federal government continue to sponsor cooperative programs, with state and local governments providing sufficient guidance to ensure conformance to national specifications and standards and thus to development of a fully integrated national information system.

and further stated that:

> We recommend that the geodetic and cartographic data bases be adequately supported, be readily accessible to all users, and even though serving different interests and needs, be made integral parts of a national mapping, charting, geodesy, surveying, and multipuspose cadastre information system.

A further statement on the procedures and standards for a multipurpose cadastre by yet another National Research Council (1983) stated that a lead agency be designated to:

> provide a structure for the formal recognition of procedures and standards for a multipurpose cadastre, as described above, and to oversee compliance with them by the federal establishment.

Other fiscal agencies in the Federal government have also recognized the need for more coordination of digital cartography and standards because of the fiscal efficiency that can be gained by better coordination of these activities. A recent General Accounting Office report (1982) recognized the duplicative nature of current computer-mapping programs in the Federal sector with a finding that in 1981 over a dozen major agencies spent in excess of $45 million on various kinds of digital mapping programs. This figure does not include any traditional hand cartography. After a thorough examination of the problem the General Accounting Office recommended that:

> the Director, OMB, issue a circular or other directive requiring the interagency coordination of computer mapping and preventing duplicative programs. The directive should create a rulemaking body to establish uniform standards for Federal computer mapping so that agencies can exchange data and the needs of map users are met at reasonable.

Their explicit recognition is intended to gain more efficiency in the use of government funds by achieving more cooperation and less overlap between agencies in this area. However one very direct implication from this recommendation is that cartographic standards must be developed to facilitate such cooperation. Just about the time that this current report was being written the Office of Management and Budget in April of this year (1983) issued a memorandum to coordinate the Federal digital cartographic program be establishing an interagency coordinating committee to oversee the process. The part of that mandate that is of interest here is that the group is charged to:

> Develop and adopt, for use by all Federal agencies, common Standards of content, format, accuracy for digital cartographic base data to increase interchangeability and enhance its potential for multiple use.

If there ever was any doubt that there is a real need to establish digital cartographic data standards that doubt has been dispelled once and for all.

Mandate for the National Committee for Digital Cartographic Data Standards

Although the tempo of recognition for the need for a national committee to recommend digital cartographic data standards has risen markedly rather recently, action to address the challenge directly has been going forward for several years. In 1980 a Memorandum of Understanding (MOU) was negotiated between the National Bureau of Standards and the U.S. Geological Survey which designated the Survey with the lead responsibility for developing, defining and maintaining data elements, and standards for earth science information systems. It should be recognized that digital cartographic standards are only one part of this overall mandate. Naturally the responsibility for the cartographic aspects of this mandate fell to the National Mapping Division of the Survey.

The National Mapping Division recognized early on that if a set of standards was to be developed that would really gain acceptance in the field of cartography, that not only the Federal agencies concerned with the result, but also the state and local agencies, the private sector and the research sector, mainly universities, must participate in this process. The National Mapping Division in all of its wisdom acted upon a recommendation from Moellering (1981) and encouraged the founding and organization of the National Committee for Digital Cartographic Data Standards to develop the standards for digital cartographic data. The Committee operates under the umbrella of the American Congress on Surveying and Mapping of which the American Cartographic Association is a member organization. With this relationship the National Committee operates under the auspices of the premier professional cartographic organization in the United States. The beauty of this arrangement is that the Committee can operate as an impartial and independent body to develop the needed digital cartographic standards in a setting that includes <u>all</u> segments of the profession.

III. ORGANIZATION AND WORK OF THE NATIONAL COMMITTEE

As a result of the establishment of the National Committee its primary goal has been defined as (Moellering, 1982):

> To provide a professional forum for all involved Federal, State, and local public agencies, private industry, and professional individuals to express their opinions, assessments, and proposals concerning digital cartographic data standards. After sufficient time for the formulation, circulation, discussion, reformulation, and comment, these proposed standards will be submitted to the U.S. Bureau of Standards to become national digital cartographic data standards.

A Steering Committee and four Working Groups (WGs) have been organized and now have been in full operation for a year. Figure 1 shows the basic organization of the Committee.

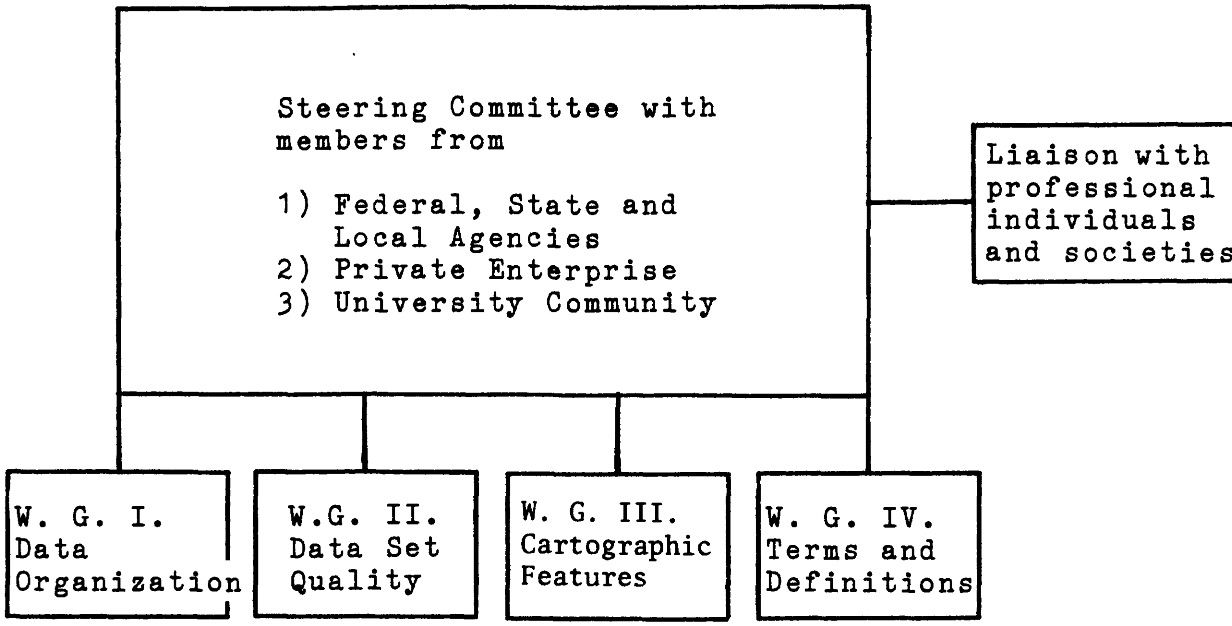

Figure 1. Basic Organization of the National Committee for Digital Cartographic Data Standards

The membership of the Committee is as follows:

Chairman: Prof. Harold Moellering, Ohio State University

Vice Chairman: Mr. Lawrence Fritz, National Ocean Survey

Members of the Steering Committee

Mr. Lawrence Fritz, National Ocean Survey
Mr. Denis Franklin, Defense Mapping Agency
Mr. Robert Edwards, Oak Ridge National Laboratory
Dr. Tim Nyerges, Private Consultant
Mr. Jack Dangermond, Environmental Systems Research Institute
Dr. John Davis, Kansas Geological Survey
Dr. Paula Hagan, Wang Laboratories
Prof. A. R. Boyle, University of Saskatchewan
Prof. Waldo Tobler, University of California
Prof. Dean Merchant, Ohio State University
Prof. Hugh Calkins, SUNY Buffalo

Working Group I, Data Organization:

Dr. Tim Nyerges, Private Consultant, Chairman
Mr. Marvin White, Bureau of the Census, Vice Chairman
Prof. A. R. Boyle, University of Saskatchewan
Dr. Paula Hagan, Wang Laboratories
Mr. Denis Franklin, Defense Mapping Agency
Mr. William Liles, Technology Service Corporation
Mr. Robin Fegeas, Geological Survey
Dr. Donna Peuquet, University of California
Mr. David Pendleton, National Ocean Survey

Working Group II, Data Set Quality:

Prof. Nicholas Chrisman, University of Wisconsin, Chairman
Mr. Fred Broome, Bureau of the Census, Vice Chairman
Prof. Dean Merchant, Ohio State University
Dr. John Davis, Kansas Geological Survey
Mr. Robert Edwards, Oak Ridge National Laboratory
Mr. George Rosenfield, Geological Survey
Mr. George Johnson, National Ocean Survey
Mr. Wallace Crisco, Bureau of Land Management
Mr. Charles Poeppelmeier, Defense Mapping Agency
Mr. John Antalovitch, Kucera and Associates
Mr. John Stout, Petroleum Information Inc.

Working Group III, Cartographic Features:

Mr. Warren Schmidt, Rand McNally Co., Chairman
Prof. Robert Rugg, Virginia Commonwealth University, Vice Chairman
Dr. Joel Morrison, Geological Survey
Mr. Robert Jacober, Air Force
Mr. Richard Hogan, National Ocean Survey
Dr. Beth Driver, Technology Service Corporation
Mr. Frederick Tamm-Daniels, Tennessee Valley Authority
Ms. Mary Clawson, IIT Research Institute

Working Group IV, Terms and Definitions:

Mr. Dean Edson, E-Quad Systems, Co-Chairman
Prof. Harold Moellering, Ohio State University, Co-Chairman
Mr. Erich Frey, National Ocean Survey, Vice Chairman
Prof. Mark Monmonier, Syracuse University
Mr. Frank Beck, Geological Survey
Dr. Carl Reed III, Autometric Inc.

Observers:

Mr. Ben Ramey, Geological Survey
Mr. Lowell Starr, Geological Survey
Mr. Henry Tom, Bureau of Standards
Mr. Roy Saltman, Bureau of Standards

Ex Officio:

Mr. Walter Robillard, President, American Congress on Surveying and Mapping
Prof. Mark Monmonier, President, American Cartographic Association
Mr. John Uehlinger, Executive Director, American Congress on Surveying and Mapping

The primary duty of the Steering Committee is to serve a policy review and coordination role to be sure that the work of the WGs is clear and comprehensive. The Steering Committee originally defined the number and scope of each Working Group and now reviews their work prior to reporting on it publicly.

The general goals of the four Working Groups are as follows:

I. Working Group on Data Organization

1) Examine cartographic data models
2) Examine cartographic data structure
3) Examine cartographic data interchange

II. Working Group on Data Set Quality

1) Fidelity of graphical data, metric and topological
2) Coding reliability
3) Update and other temporal information
4) Lineage of a data set
5) Checking procedures by the producer to verify quality

III. Working Group on Cartographic Features

1) Define feature classes
2) Define structure and levels of classes
3) Define feature codes

IV. Working Group on Terms and Definitions

1) Collect new terms defined by working groups
2) Define other new terms

Conceptual Background

The conceptual background for this effort has been defined in detail in an article in NCDCDS Committee Report No. 1 (Moellering, 1982) but shall be briefly summarized here.

> The conceptual milieu in which the committee is operating has also expanded dramatically in recent years as noted in the earlier paper. The concepts of real and virtual maps greatly clarify the situation of the new digital cartographic products and how they relate to the more conventional products (Moellering, 1980). Transformations between real virtual maps define most important operations in cartography and have been an

interesting concept for the design of modern cartographic systems. Nyerges (1980) has devised the notions of deep and surface structure as they apply to cartographic information and has shown that surface structure representations of cartographic information are real and virtual type I maps while cartographic deep structure is usually represented in the digital domain by type III virtual maps. It is also possible to look at these standards efforts in terms of deep and surface structure. Surface structure is the graphic representation of cartographic information such as a conventional map or CRT display. Over the years many principles have been defined for cartographic design which must be followed if one is to have an effective map. However, the deep structure, that area of spatial relationships between cartographic elements of cartographic information which are not graphic, is where must of the digital information resides which is stored in modern cartographic data bases. In essence, the primary task of this committee is to bring conceptual order to the area of deep structure in digital cartography.

IV. ISSUES FACING THE DEVELOPMENT OF DIGITAL CARTOGRAPHIC STANDARDS

In order to develop effective digital cartographic standards the most efficient approach is to follow the strategy that is the same as one would the solving of any other scientific problem. Therefore one begins with the general considerations and progressively works down to the specific detailed problems and then backs up to the general problem while at the same time achieving an integrated solution to the problem. As a result the solution process has been conceptualized into five basic cycles of work:

1) Define the fundamental issues involved

2) Define the alternatives to the problem

3) Formulate interim proposed standards

4) Reformulate interim proposed standards

5) Generate final proposed standards.

At the end of each cycle a report is written by the committee and circulated to the profession for thought, reflection and comment. Comments received by the Committee from concerned professionals will be integrated into the process immediately. It should be fairly clear that this incremental process is designed to minimize contrasting opinions at the end by including comments immediately into the process.

If one is to achieve an efficient solution to the problem of digital cartographic standards, one must begin by thoroughly addressing the issues. It is only until one can succinctly enumerate the issues, or in terms of the philosophy of science state the hypotheses to be tested, can one begin to actually solve the problem(s) at hand. The balance of this report deals with a discussion of the issues facing the Committee and the profession. There are four classes of issues, each set directed towards a specific Working Group. Each WG report examines the issues in its area and specifies the scope of the problem that the WG will attempt to solve. As Chairman of the Committee, the author urges each reader to carefully consider the issues as they are discussed. These statements will serve as a blueprint for the future work of the Committee. The draft report of each Working Group presented here have been previously circulated through the Steering Committee for review and comment. In general, the members of the Steering Committee concur with the thrust of these statement. Now it is the turn of the members of the profession to comment on these statements.

V. FUTURE WORK OF THE COMMITTEE

The WGs have now begun on the second cycle of the process, that of examining the available alternatives. The fundamental thrust of this effort is to examine and explicitly appraise the advantages and disadvantages of possible alternative solutions of these problems as stated in the issues being considered. In the early part of 1984 the Committee plans to issue Report No. 4 on the available alternatives. It is further planned that this report will be issued in time that members of the profession will be able to digest and consider the report and then participate in open hearings at the Spring 1984 meeting of the American Congress on Surveying and Mapping in Washington, D.C. All interested individuals are invited to attend.

The question also arises as to the final disposition of the standards developed by this Committee. One will recall from the earlier discussion at the outset of this paper that the Committee is operating in response to a Memorandum of Understanding between the National Bureau of Standards and the U.S. Geological Survey. Therefore the proper disposition of the standards developed here is to submit them back to the Bureau of Standards via the USGS channel (National Bureau of Standards, 1983). The Bureau of Standards will then consider the submitted standards with the intent of having them become Federal Information Processing Standards (FIPS).

VI. SUMMARY AND CONCLUSION

The need for digital cartographic data standards has been recognized for more than a decade. The National Committee was founded in response to that need. The Committee has now finished its first cycle of work, that of defining the issues. Work on the second cycle of examining the alternatives is progressing.

VII. REFERENCES

Federal Mapping Task Force, 1973, Report of the Federal Mapping Task Force on Mapping, Charting, Geodesy and Surveying, Office of Management and Budget, Washington, U.S. Government Printing Office Report #2401-02426.

General Accounting Office, 1982, Duplicative Federal Computer-Mapping Programs: A Growing Problem, Comptroller General of the United States Report #GAO/RCED-83-19, Gaithersberg, MD: General Accounting Office.

Moellering, Harold, 1980, "Strategies of Real-time Cartography" Cartographic Journal, 17(1), pp. 12-15.

______________, 1981, A Proposed Approach to Establish a Set of National Digital Cartographic Data Standards, Report to the National Mapping Division, U.S. Geological Survey.

______________, 1982, "The Challenge of Establishing a Set of National Digital Cartographic Data Standards for the United States," Proc. American Congress on Surveying and Mapping 42nd Annual Spring Meeting, Denver, pp. 201-212.

National Bureau of Standards, 1983, private correspondence.

National Research Council, 1980, Need for a Multipurpose Cadastre, Washington: National Academy Press.

______________, 1981, Federal Surveying and Mapping: An Organizational Review, Washington: National Academy Press.

______________, 1983, Procedures and Standards for a Multipurpose Cadastre, Washington: National Academy Press.

Nyerges, Timothy, 1980, Modeling the Structure of Cartographic Information for Query Processing, unpublished Ph.D. dissertation, Department of Geography, Ohio State University.

Office of Management and Budget, 1983, "Coordination of Federal Digital Cartographic Data Programs," Memorandum of April 4.

Issues in Digital Cartographic Data Standards
A Progress Report
Working Group I

Prepared by
Timothy Nyerges and
Marvin White

The Data Organization Working Group is composed of the following members:

Dr. A. Raymond Boyle	- University of Saskatchewan
Mr. Robin Feagas	- U.S. Geological Survey
Mr. Denis Franklin	- Defense Mapping Agency
Dr. Paula Hagan	- Wang Labs
Mr. William Liles	- Technology Services Corporation
Dr. Timothy Nyerges	- Consultant
Mr. David Pendleton	- National Ocean Survey
Dr. Donna Peuquet	- U.C. Santa Barbara
Dr. Carl Reed III	- Autometric, Inc. (TERMS Working group representative)
Mr. Marvin White	- U.S. Bureau of the Census

INTRODUCTION

Purpose

The Data Organization Working Group received its charter from the National Committee for Digital Cartographic Standards (NCDCDS) Steering Committee in June, 1982. Its purpose is to examine issues, identify alternatives and propose standards for digital cartographic data organization.

History of Meetings and Communication

The first meeting of the full membership of the working group was in August, 1982 at the Auto-Carto V meetings, and the second meeting was held at the Spring 1983 ACSM meetings. Several members wrote postion papers to identify issues. Some of these papers were discussed at the initial meeting, and later papers were circulated among the group members and discussed at the second meeting of the group. These papers and discussions formed the basis for this paper and corresponding presentation at the Spring, 1983 ACSM meetings.

Overview and Motivation

The impetus for our committee's work is cartographic data interchange. It is plain to everyone with some experience in exchanging digital cartographic data that impediments to data interchange are not simple matters of format, rather they stem from fundamental differences in recording cartographic data. An illustrative example is transforming a raster image of a map into a polygon boundary encoding. This job is formidable; it demands a deep understanding of exactly what is recorded in the two forms and how they relate to each other. Consequently, our first task is to find or invent a model (of the Earth) that is general enough to encompass the implicit models of the various map encoding methods. We need a model for our example that permits us to capture both raster and boundary encodings, and only then can we prescribe standards of interchange. Of course, there are other methods of map encoding and our model must accomodate them as well.

Major Issues

The Data Organization Working Group identified issues in three major areas of digital cartographic data organization. These areas are:

1. Terminology
2. Modeling
3. Data Interchange

These areas do overlap with each other and with the areas of concern to other working groups. The distinctions are made to help focus our deliberations.

Rapid development by many diverse workers in the field has given rise to a large technical vocabulary having several different words with very similar meanings and some words with more than one distinct meaning. This tends to confuse rather than clarify issues and confound our deliberations. As part of our effort, the working group members will try to eliminate some of the confusion, by focusing on generic elements in spatial modeling and data interchange. Because terminology is so important in communication of ideas, we have decided to make the topic a major issue in itself. This work will naturally be conducted in cooperation with the Terms working group.

The second major issue is spatial modeling. The first problem we encountered here was terminology; the group could not reach agreement on the meaning of "data model," so we speak only of modeling. We decided to undertake a review of the conceptual aspects of data organization existing in spatial models based on cartographic and mathematical theories. The topic of cartographic data structure is to be subsumed under this topic. Focusing on the conceptual aspects of data organization, rather than data representation, will provide a theoretical base from which we can identify spatial model completeness. That is, what models are used for specific problems, and what problems can be solved by what models. The theories are the source of much of the terminology of the field, so our investigations of spatial modeling will also contribute to our work on terminology.

A theoretical understanding of spatial models will also help clarify the issues of data interchange, the issue that motivates the entire effort. The data interchange issue is third on the list but primary in importance. We have agreed to identify past and potential problems of data interchange. In addition, we will describe a generic data interchange varying by application areas and data components.

TERMINOLOGY

Terminology for data organization is one of the major issues before us. We will try to consider all fundamental digital cartographic data organization terms with the help of our representative from the terminology working group. Those of special importance will be mentioned later in this paper under the spatial modeling and data interchange issues.

Importance of Terminology

The terminology issue is mentioned first because it affects all topics of concern in this process of standards development. The task of identifying conventional terminology has both disadvantages and advantages. The disadvantages concern the limitations which can be forced upon topics by viewing them under the guise of certain terminology. The advantages concern the clarification which can result from identifying the terms which have similar and dissimilar meanings.

We would like to discuss topics utilizing generic terminology. Such terminology is not always easily identified. White (1979, 1982) proposes we use terminology rooted in mathematical theory; for this terminology facilitates clarity through generality of concept. However, even though our terminology may be clearly defined, many of our current problems with data organization are a result of the particular viewpoint with which we approach these problems. Peuquet recommended classifying terms as: 1) forming part of our framework for discussion, 2) common but possibly ambiguous, and 3) so ambiguous that they should be avoided.

Views of a Topic

Two fundamentally different approaches to formulating a spatial model of reality can have different affects on the form and substance of that model. Two such approaches are a practical problem approach and an abstract phenomena approach (Hagan, 1982; Nyerges, 1982).

The practical problem approach is an inductive method of specifying the nature of a problem through an identification of entities concerning a given problem on the basis of experience. This approach is also called a view modeling approach, because different views are aggregated to create a general model. The model, however, is only as thorough as the views used to describe spatial oriented problems. This process tends to focus problem solutions toward representation and implementation.

The abstract problem approach is a method of deducing the inherent characteristics of phenomena. This approach is also called phenomena modeling because an attempt is made to model the form of phenomena without regard to particular problem orientation. Such an approach tends to start at high levels of abstraction and proceeds to more concrete decompositions when creating conceptual models.

Although the two approaches can be distinguished on a theoretical level, it is difficult to say that one can adhere to either approach when elucidating the nature of spatial models. It is the intention of the Data Organization Working Group to keep both approaches in mind.

Examples of Terminology

Since basic definitions of terms are usually taken as assumed knowledge, it is important that correct meanings result from the use of these terms. The cartographic/geographic lexical framework specified using Backus-Naur Form in Appendix A of the DOE Interlaboratory Working Group on Data Exchange (1978) is one good example of a well thought-out framework for communication. The terms are clearly defined in regard to each other.

In addition, James Corbett's (1979) and Marvin White's (1982) use of n-cell terminology is another example of a consistent framework for communicating the basic elements of digital maps. Such generic approaches as these are critical if we are to eliminate ambiguity in the discussions of modeling and interchange of digital data.

MODELING

Spatial modeling is the second major issue identified by the Data Organization Working Group. Modeling requirements for practical problems are considered before discussing the formulation of an abstract model. A discussion of representation versus modeling follows and the section ends with a discussion of modeling completeness.

A Framework for Data Organization

We seek a comprehensive framework to conduct our work and ultimately to specify standard methods for organizing data to facilitate interchange. Currently, the framework in which we are working has 5 levels related to the levels described by Nyerges (1980):

1) Reality - (Our perception of) the Earth. We are not thinking of modeling maps, rather we take digital cartographic data as representing features on the Earth, regardless of the immediate source.

2) Theory - A coherent explanation of cartography. At least one member of the group refers to formalized theories.

3) Guide to Interpreting the theory - These are the abstractions in passing from reality to a model. For example at a small scale a road is a linear feature, but at large scale it occupies a definite area.

4) Model - This is the interpretation of the theory. Some members regard models as very specific, i.e., a particular file is a model of a piece of the Earth just as a model airplane is a model of a particular large airplane. Others take models to be less specific and speak of data models, etc. The matter is still under discussion.

5) Machine Encoding - The actual representation in a computer taking account of word size, architecture, etc. This is the stuff to be interchanged.

Modeling Requirements in Practice

As mentioned in the terminology section, two approaches help characterize spatial data modeling. These two approaches are practial problems and abstarct models.

We can identify practical modeling requirements through a practical problem approach (Hagan, 1982). This involves identifying and defining classes of cartographic data processing from which underlying principles can be extracted. Then we can define modeling characteristics and modeling problems for each class of processing applications. As a third step we can identify and define operations to be performed for each class of applications. Finally, we can define a user method for expressing those operations to be performed. These steps will clarify the practical or application-oriented side of the modeling issue.

Formulating an Abstract Model

Data model in the computer science literature is defined as a generalized data structure having a set of operators for processing basic structures and using integrity constraints to check data integrity (Date, 1982). Thus, we describe a data model as having three fundamental components: structure, operators and constraints. The structure is the data organization for explicitly stored data and relationships, while the operators makeup the component for deriving further information. The hierarchical and network data models rely quite heavily on explicitly stored data and relationships. The relational model relys quite heavily on operators to be used in the derivation of information from stored data. Both models utilized rules to check data integrity.

The term 'data model' has not been widely used in the cartographic data structures literature. However, the term model has been used quite heavily for some time, since the introduction of the spatial organization paradigm in Geography in the 1950's. Although it has been shown that any one of the three traditional data models can be used for cartographic data structures, it has also been shown that none of them are generally suitable. This is the case because cartographic data structures require the speed of hierachical structures, the modeling capability of network structures and the flexibility of relational structures. However, we should be able to identify application areas suitable for each of the three models (Pendleton, 1983).

We need to clarify our use of model. Are we to use the term data model to refer to a model of a map or should we use the term to refer to a conception of geographic space (Peuquet, 1983)? The former may be considered a model of a model, whereas the latter is a model of reality. In the past, cartographic data structures have been developed with both in mind.

Clearly, data model should be used to connote something more general than a data structure. Perhaps this is why the ANSI/SPARC group on database management agreed to replace the term "data model" with "conceptual model" for characterizing the general model of a database. Unfortunately, the group failed to devise a generalized data model that could act as a covering set for all three traditional models. Thus, the term conceptual model is still somewhat conceptual.

We are interested in identifying fundamental characteristics of data. Whether we categorize characteristics such as topology and geometry by module or component or class, they represent the generic aspects of the digital data for which we are to propose standards. Regardless of whether these characteristics are implicit or explicit for certain structures, all possible categories of such generic characteristics need to be enumerated. Enumeration of these basic categories should help us distinguish between modeling characteristics and data representation characteristics.

Basic Elements in Representations vs. Modeling

Distinguishing between data modeling and data representation will permit us to better understand the topic of data organization. This elucidation has already begun, for example, in the discussions by James Corbett (1979) and Marvin White

(1979) concerning the mathematical theory of 2D maps. Such expositions will need to continue in the future in the more general context of spatial data processing. Numerous concepts have been proposed as basic elements of cartographic models; to name a few:

1. Geometric topology (0- 1- and 2-cells): Corbett (1979) and White (1979)
2. Basic Elememts: DOE (1978)
3. Logical and Physical Abstract Data Types: Burton (1979)
4. Relations: Shapiro (1979)
5. Primitive and Compound Objects: Yqungman (1977)
6. Class, object, attribute, link: Bouille (1978)

The above elements tend to focus on a mixture of modeling and data representation. There has been considerable misunderstanding between the two topics, especially as related to cartographic data structures. Data representation has received considerable attention because of a desire to develop digital cartographic pragmatism. The overemphasis on data structure/data representation pragmatism has tended to lead us away from the more general problem of modeling geographic space, or entities in geographic space (depending on ones conception of space). A focus on data model completeness rather than data structure generality can show us the need for renewed emphasis on modeling. As our difficulties in terminology reveal, we cannot avoid theories and models--they tell us what our terms and data mean.

Model Completeness

Relational completeness according to Codd (1972) is the measure of the selective power of a query language. The relational algebra and relational calculus are said to have equivalent measures of relational completeness. The term completeness, therefore, concerns the ability to retrieve information in the database, whether stored as records or relations.

Here we define data model completeness which is different from relational completeness. Our interest is in being thorough about all data to be stored or derived in a database. Thus completely describing cartographic phenomena in digital form rather than the completeness of the query language for retrieval is of primary concern. One of the methods for measuring such completeness is to compare practical database requirements with what the abstract model will allow. Another might be to perform model to model algebraic mappings to uncover structures common to them. This question of completeness becomes very important when trying to interchange data. We would want to know what information is missing from one data model (data structure) that may or may not be able to be generated in the context of another.

All questions of data are formulated, in a context. Reliability of answers is only valid with respect to context.

A Proposed Model

White (1982) proposed adopting Corbett's (1979) model as the general purpose model within which to interpret the various coding methods and specify interchange standards, at least for 2-dimensional maps. This means that interchange standards would be specified in terms of the topological theory of maps (using words like "0-cell" and "incidence"). The working group has not agreed that this model should be adopted nor even that it encompasses all methods of encoding. We present a sketch of it below only to report progress and indicate the general form of our final product.

In the topological model, there are only three different kinds of objects, which are the 0-cells, 1-cells, and 2-cells. In general an n-cell is an n-dimensional object; so a 0-cell is a point; a 1-cell, a linear feature; and a 2-cell, an areal feature. A map is regarded as a jig-saw puzzle comprising 0-cells, 1-cells, and 2-cells. The n-cell terminology is taken from the topological theory originated by Henri Poincare circa 1900 and has the advantage of not being colored in meaning by special uses in cartography.

The objects (n-cells) are related to one another only by incidence. Pieces that fit together in the jig-saw puzzle are incident and those not touching are not immediately related. (Of course, they may be related by having some of the same neighboring pieces but these are implied relations and can be devised.) The objects are related to Euclidean 3-dimensional space by coordinates and shape. 0-cells have coordinates (x, y, z), 1-cells have shape (straight, circular, passing through specified points, etc.), and 2-cells have shape (specified by contours or profiles or the like). This completes the geometrical model.

The n-cells have other (non-geometrical) attributes such as name. A 1-cell representing a river would have a name as well as an indication that it is a river. A 2-cell might have several geographic codes and jurisdiction names. There is no essential limit to the associated attributes in the model. Using this model we can represent a map although it may be recorded digitally in two different ways.

DATA INTERCHANGE

Data Interchange is our third issue and our most important in regard to practical results.

Review of Formats

Thus far, eight data interchange oriented documents have been identified for review:

- Australian Feature Coded Digital Mapping Data Standard
- Canadian EDP Standards Applied to Digital Topographic Data
- Computer Assisted Mapping and Records Activity System
- DOE Interlaboratory Working Group on Data Exchange
- DMA Standard Format for Linear Digital Data
- Graphics Standards Planning Committee Appendix Report on Metafiles
- Initial Graphics Exchange Specification
- Intergraph Standard Interchange Format

A number of topics are of concern when we compare and contrast the content of these documents. Each is being examined for its general utility. This utility is measured in terms of its general applicability to all topics in digital cartography. Thus, document, application orientation is being identified, as well as potential applications.

The data representation/modeling aspect of the documents is especially important, both for logical and physical representation. Basic data elements, data structures and coordinate systems are especially important logical characteristics. It is crucial that we focus special attention on data representation in the context of cartographic application orientation. Enumerating the critical characteristics of all data representations can be extremely helpful in our assessment of alternatives.

The physical characteristics for data representation are a very important part of digital data transfer. Record formats and hardware orientation mentioned or implied in the documents is of special interest. The more physical bias, the less well suited the format for general use.

We are trying to identify and enumerate problems currently encountered with the formats. In addition, we are conjecturing about problems that may eventually surface.

A Generic Data Interchange

From the review of these documents we will have a better understanding of the alternatives for digital data interchange. We know that a generic data interchange approach is the only approach that may satisfy all applications. This approach may involve a statement of the data modules or data components in a generic fashion. Such an approach may also involve a thorough investigation of meta-data for self-describing files. Meta-data would announce what the data is like internally without being forced to result to external information.

Using the topological model outlined above for data interchange, we would proceed as follows for incorporating polygon outlines (say county boundaries) in a triangulated Digital Terrain Model. The triangulation is easily acommodated in the model: the triangles are 2-cells; the boundary lines separating two triangles, 1-cells; and the vertices, 0-cells. The county boundaries are a little more difficult. The counties are 2-cells, but the boundary traces are not 1-cells, because the common boundary between two counties is given in two parts: one for one county and one for the other. Once the boundaries are matched (not a small task), the 1-cells and 0-cells can be identified. Finally, the two maps are combined, and this is no small task either: The n-cells common to both must be identified and those that must be partitioned to correspond are next partitioned, and then the data can be exchanged. A model is essential in doing such an exchange.

Interchange Alternatives

Finally, at issue here is the basic nature of interchange itself. We will look at the immediate advantages of software oriented approaches with exchange via magnetic tape, as well as speculate on hardware oriented direct links (Boyle, 1982).

SUMMARY

The Data Organization Working Group has been actively pursuing the identification of basic issues for digital cartographic data organization since August, 1982. The members on the working group represent views from government agencies, universities and private industry. Through the diverse backgrounds of members of the working group we hope to establish a sound conceptual basis for discussing cartographic data organization in regard to terminology, spatial models, and data interchange. The issue concerning terminology was introduced first because of the importance of clarifying topics. This is crucial to the exchange of ideas, e.g., many of the terms used in this paper are not widely used in digital cartography. We want the professional cartographic community to be aware that new data organization terms must be integrated into our vocabulary to keep us abreast of advances in other professions as well as our own.

The second major issue is spatial models. An understanding of the differences among structures, models and representations is critical because of the way they have clouded the focus of cartographic theory development. A sound theoretical basis is the key to understanding what we are trying to standardize and what we are not trying to standardize.

The third major issue is data interchange. Clearly, this is the area of immediate practical concern. If we do not have a sound theoretical base from which to approach this topic, then we will fail at our task. A review of the eight identified graphics data exchange oriented documents shows a diversity of ways that graphics data can potentially be exchanged. However, some common threads also weave their waythrough these documents. It is our task to identify these differences and commonalities so we may be able to readily assess the alternatives for data transfer. In addition, we will then be in a better position to propose a national standard for digital cartographic data transfer.

CONCLUSIONS

We are getting the gaps in our knowledge concerning data organization into focus. We see four specific areas wanting continued research by the professional community:

1) A comprehensive theory of computer assisted cartography that permits such activities as interchanging data among users with very different digital data, e.g., DIME files and raster images;

2) A better understanding of abstraction, especially as it relates to scale.

3) A better understanding of the variety of users' viewpoints and needs (how can a forest manager even discuss data interchange with an urban planner?);

4) Transfer of large volumes of data--even if we had standard theories, models, and formats, how can map files be feasibly transferred?

For much of digital cartographic history, a desire for digital cartographic data structure pragmatism has resulted in a vast amount of disjointed research; with few papers considering spatial theory. The structures we have been developing are limiting in terms of spatial meaning. This is in part due to our lack of understanding of the more general problem which involves the modeling of geographic space or entities in geographic space, depending on ones conception of space. A few models have been proposed, but a comprehensive theory has not been formulated. What we need in the near future is more emphasis on spatial modeling in a theoretical context. Anything a spatial scientist can do to help bring this about would clearly be beneficial to all the information professions.

Our next step in the cycle of developing standards is to identify alternative strategies for the basic issues identified to date. This involves specifying working definitions for terms we use, describing the nature of a spatial model, and weighing the alternative methods for data interchange. Each one of these is a large task onto itself.

What can we expect to accomplish in the next year? A number of working definitions for new terminology will be proposed. It is doubtful that a general spatial model will be proposed; however, advances have been known to occur. We will probably be able to propose a set of generic data characteristics that can be included in an interchange format generally applicable to digital cartographic data.

Lastly, we conclude this discussion by calling for professional support in the form of constructive criticism. If we have overlooked materials and/or topics, please let us know by way of the NCDCDS communication channels.

References Cited

American Public Works Association, Utility Location and Coordination Council (1979). "File Format for Data Exchange Between Graphic Databases," Vol. 3 of Computer Assisted Mapping and Records Activity System. January, 1979.

Bouillé, F (1978). "Structural Cartographic Data and Spatial Processes with the Hypergraph-Based Data Structure." Harvard Papers on Geographic Information Systems Vol. 5, G. Dutton, ed. Cambridge, MA: Harvard University Laboratory for Computer Graphics and Spatial Analysis.

Boyle, A.R. (1982). "Problems with Interconnects of Systems," unpublished position paper on Data Organization Working Group Issues. August, 1982.

Burton, W. (1979). "Logical and Physical Data Types in Geographical Information Systems," Geo-Processing 1:167-181.

Canadian Council on Surveying and Mapping (1982). Canadian EDP Standards Applied to Digital Topographic Data, Energy, Mines and Resources, Canada. April, 1982.

Codd, E.F. (1972). "Relational Completeness of Data Base Sublanguages." In Data Base Systems, Courant Computer Science Symposia Series, Vol. 6. Prentice-Hall.

Corbett, J.P. (1979). Topological Principles in Cartography, U.S. Bureau of the Census, Washington, D.C.

Data Organization Working Group I. (1982). "Report on Activities of Working Group I, Data Set Organization" in Selected Papers from the Session on Digital Cartography - Planning and Standards at AUTO Carto V. Edited by H. Moellering. National Committee for Digital Cartographic Data Standards.

Date, C.J. (1982). "A Formal Definition of the Relational Model," ACM SIGMOD Record, Vol. 13 Number 1, September, 1982, pp. 18-29.

Defense Mapping Agency Aeorspace Cartography Department (1981). "DMA Standard Format for Linear Digial Data," Defense Mapping Agency Aerospace Center. November, 1981.

Graphics Standard Planning Committee (1979). "Graphics Metafile Description," Computer Graphics Quarterly, 1979.

Hagan, P. (1982). "Digital Cartographic Data Standards Data Organization Working Group Suggested Goals," unpublished paper. August, 1982.

Intergraph Corporation (1980). Graphics Standard Interchange Format. November, 1980.

Interlaboratory Working Group on Data Exchange (1978). Geographic Exchange Standard and Primer. U.S. Department of Energy. October, 1978.

National Bureau of Standards (1980). Initial Graphics Exchange Specification. U.S. Department of Commerce, now published as ANSI Standard Y14.26M.

Nyerges, T.L. (1980). Modeling the Structure of Cartographic Information for Querying Processing. Unpublished Ph.D. dissertation, Department of Geography, The Ohio State University.

Nyerges, T.L. (1982). "Position Paper on Data Organization Issues," unpublished paper. August, 1982.

Pendleton, D. (1983). "Phase I: Defining the Issues," unpublished position paper for National Digital Cartographic Data Standards Committee Working Group on Data Organization Unpublished paper, January, 1983.

Peuquet, D. (1983). "Position Paper for National Digital Cartographic Data Standards Concerning Data Organization Working Group Issues." Unpublished paper, February, 1983.

Shapiro, L.G. and R.M. Haralick (1979). "A Spatial Data Structure," Technical Report #CS 79005-R, Department of Computer Science, Virginia Polytechnic Institute and State University.

Standards Association of Australia (1981). Interchange of Feature Coded Digital Mapping Data Australian Standard 2481-1981. Standards Association of Australia Standards House.

White, M. (1979). "A Survey of the Mathematics of Maps," Auto Carto IV, Reston VA.

White, M. (1982). "Comprehensive Mathematical Model of a Map," unpublished position paper on Data Organization Working Group Issues. September, 1982.

Youngman, C. (1978). "A Linguistic Approach to Map Description." Harvard Papers on Geographic Information Systems Vol. 4, G. Dutton, ed. Cambridge, MA: Harvard University Laboratory for Computer Graphics and Spatial Analysis.

(p. 23 follows)

Issues in Digital Cartographic Quality Standards
A Progress Report

Working Group II, NCDCDS
prepared by N. Chrisman

Working Group II on Data Set Quality is composed of:

Nicholas Chrisman (chair)	University of Wisconsin
Fredrick Broome (vice-chair)	U.S. Bureau of the Census
John Antalovich	Kucera and Associates
Wallace Crisco	Bureau of Land Management
Robert Edwards	Oak Ridge National Laboratory
George Johnson	National Ocean Survey
Dean Merchant	Ohio State University
Charles Poeppelmeier	Defense Mapping Agency
George Rosenfield	Geological Survey
John Stout	Petroleum Information Inc.

The working group met on March 13 at the Washington Hilton and was able to complete a general list of issues which must be addressed to create standards for data quality.

General nature of data quality standards

The quality characteristics of digital cartographic data are a consequence of its fitness for use. Clearly, different uses demand different forms of quality, so it is counterproductive to demand rigid thresholds as standards. Characteristics of the information should be set forth in a quality report by the producer so that potential users will be able to evaluate the benefits and limitations of a digital source relative to a particular set of requirements. To a large degree, the projected quality standard is a "Truth in Labeling" effort, not a rigid specification of procedures, accuracies and so on. Much of the information in a quality report should be known by cartographic producers, but there has been no established mechanism to communicate this information to users.

In the coming digital age, the ability to misjudge quality and abuse cartographic information is dramatically increased. By supplanting traditional graphic forms of geographic information, digital cartography calls for a thorough reexamination of cartographic production practices from source material to products and back through the revision cycle. Quality standards must be carefully designed not to hinder future developments, instead they should encourage a better fit between data sources and application needs.

1. Lineage report and assessment of suitability

The fundamental issue in understanding a digital data product is to reach behind that product to probe the processes which produced it. One of the best means for a user to understand the potential information is to understand the motivations and objectives of the producing group. While such a statement may be broader than the quality issue alone, a written description of the objectives for capturing the digital data would provide an aid in assessing fitness to a user's application.

Beyond overall objectives, the written description should continue into substantial technical detail in describing the "lineage" of the digital data. By lineage we mean the narrative of the processes and transformations which started from source material and resulted in the digital product. Some production details may be lost in the murk of the past, but the standard should promote as complete disclosure as possible. Lineage information should provide a framework for much of the other facets of quality assessment.

A lineage statement defines the product and how it was arrived at through the technical processes. In addition there may be a need to discuss the suitability of these procedures as applied to the information encoded. This statement should not intend to defend a particular theory of map encoding, but rather discuss the rationale for applying it to the given problem. This discussion should also include a producer's evaluation of the aptness of the source material to the particular application. For example, an agency might be forced to choose between airphoto coverage that is recent but of the wrong season, or more out-of-date but more interpretable. A lineage statement, augmented by these evaluations of suitability, should permit informed judgement of the relevance of the data to the particular problems of the user's different application.

2. Specifying and testing data quality

Some of the most important issues relate to the quantitative evaluation of various aspects of data quality. This section will discuss the subject matter which should be subjected to tests, while the next ones will deal with test procedures and their results

The contents of a digital cartographic data base can be decomposed according to different schemes, and the logic of a given application may lead to a particular scheme seeming "natural". In the general case it is more complicated, but for the purposes of this discussion four general components have been identified: logical consistency and completeness, space, attributes, and time.

2.1 Logical consistency and completeness

An essential ingredient in assessing the suitability of a digital cartographic data base is a knowledge of its logical properties. This issue cuts across all components of the information and deserves priority treatment. In the computer science literature, integrity refers to validity of a single, isolated value, while consistency refers to properties of two or more values. Both properties must be evaluated to determine quality and fitness for use. Some of the errors in logical structure will be errors of commission - inconsistencies, while others may be errors of ommission - incompleteness.

One of the major elements of logical quality involves the data model and data structure employed to encode the information. Experience with digital cartography demonstrates that there is a decisive difference between formulating a theoretical data structure and ensuring that it is faithfully implemented. For example, the U.S. Census Bureau has adopted a rather complete topological encoding for its DIME files. This structure has the ability to detect many logical inconsistencies, but the Bureau - for good practical reasons - can not afford to remove all errors from their digital files. With other simpler data structures it may not even be possible to detect inconsistencies automatically. In either case, the user needs to know the integrity of the data coding through standard tests and procedures.

Other components of logical consistency relate to the comparison of the digital product to the ground. Any map is a simplified and generalized abstraction of the reality of the earth's surface; the rules for this abstraction should be consistently applied. With point and line features there are rules for inclusion based on proximity to other features or importance. A user would like to know what portion of the hydrographic network is actually encoded, and what criterion was used for selection. The old rules for the graphic product may not apply to digital data bases.

Similarly, areal maps have logical properties which must be examined. Was the classification scheme exhaustive? Were minimum mapping units or widths applied, and how consistently? How were zones of mixed type handled?

All these concerns are crucial to inform a user of the fitness of a digital data file to a particular application. It is important to note that these concerns of logical structure are more fundamental than any statement of accuracy. Some form of specifying and testing logical properties should be developed as a part of a data standard. Producers may know important portions of this information, the task is largely how to provide a comprehensible communication of this information.

2.2 Space

All digital cartographic information contains a spatial component, either explicitly as in the form of coordinates, or implicitly as in a raster format. The spatial reference should be subjected to a set of standardized tests to evaluate fitness for use.

Positional references should be related to some coordinate system, hopefully a standardized or easily generated one. The producer should specify in the lineage statement the procedures used to derive these coordinates and used to link the digital data to established control points. For instance, the nature of "geometric correction" of remotely sensed imagery should be explained.

In cartographic practice there is a distinction between relative and absolute accuracy, but the presentation of digital data, often as absolute coordinates, might obscure such a distinction. It would be easy to jump from a desire for quality in spatial measurements to a study of positional accuracy. The fitness for a particular use might not fit precisely into this narrow focus, however. Applications of digital data might be concerned with secondary spatial measures derived from coordinates. Some of these measures, such as equidistance, azimuthality, conformality, et cetera can be correctly preserved even when absolute accuracy is imperfect. Of course, a highly accurate and precise representation will satisfy the need for derived quantities, but some needs could be served in other ways.

Digital methods are often used to overlay diverse sources of information on the basis of coordinates. In such applications the absolute positional accuracies of features become quite important. Files generated for relative positions should be clearly labelled so that overlay operations are not improperly applied.

Highly developed methods exist to establish positional accuracy in the sciences of surveying, geodesy and photogrammetry. It is less common to evaluate positional accuracy after all the phases in the cartographic production process, but essentially the same techniques apply. The nature of cartographic features is more complex than the "well-defined" points usually treated in other disciplines, but there are mathematical procedures to develop appropriate models. In developing standards to describe positional accuracy, this working group will rely heavily on parallel standards efforts in other disciplines.

2.3 Attributes

Just as all digital cartographic data involves space, it also involves a thematic or "attribute" component. The nature of an attribute ranges broadly from arbitrary identifiers through simple feature codes to measured phenomena. Due to this divergence, it will be harder to make sweeping standards for this component.

In some other countries' draft standards, the fidelity of attribute information is given much less treatment than positional error. This working group does not share this attitude. Errors in attributes are at least as likely as errors in positions. Both have an impact on fitness for use.

The statistical treatment of errors in attributes is an emerging field, most prominent in applications of remote sensing. Classification accuracy has many similarities to statistical treatment of medical diagnosis, so the field has related developments to draw upon. Attribute information can be affected by improper design of the classification scheme, misinterpretation, miscoding and other errors. Just as positional quality can be assessed by tests, attribute accuracy can be described by a different kind of test.

2.4 Time

The last component of cartographic information is the temporal one, often called currency. Most maps represent a specific date or a period. A legend may read "Compiled from 1956 air photography, field checked 1959". In such a simple case there is little complexity to the temporal component. However, temporal integrity can become questionable as a map is partially revised. The future digital age will make it easier to change cartographic information, so the issue of temporal reference will become more important. In cases where the world is changing, inaccurate temporal information can create the impression of positional and attribute errors. In reality, all these components must be linked in a comprehensive system for assessing fitness for use (quality).

3. Developing testing procedures

A major part of this working group's efforts will go into the development of standardized procedures to evaluate the quality of digital data. Some tests will examine a particular component (positional accuracy, fidelity to data model, et cetera), while other tests may involve combinations (time compounded with classification). Despite these differences, tests will fall into categories depending or the degree of rigor the test applies.

The lineage report (discussed in section 1 above) provides the basis for quality testing, but it does not constitute a test itself. The lineage report will cover a number of topics; each one could be upgraded from a description by the availability of quantitative estimates. These figures should derive from established, standardized testing procedures. In order of increasing rigor, the categories of test are deductive estimates, internal evidence, source comparison, and independent evidence.

3.1 Deductive estimates

The simplest category of quantitative measurement comes from a deductive approach. Strictly speaking, deduction does not involve a test (or inductive) procedure applied to the information at hand. In many cases, however, it will be sufficient to refer to a quantitative estimate established in the professional literature. For example, it may not be necessary to perform an accuracy study on a given surveying instrument using the particular map sheet. Reference to a distinct test procedure should be adequate, although there is a risk of different circumstances leading to different results. Deductive estimates may be particularly appropriate for a producer to refer to exhaustive tests performed on a sampled basis - a few sheets from a whole coverage.

Deductive estimates should be carefully documented in their assumptions and mathematical formulation. For instance, it is possible to deduce overall error when error in the component processes are established. Usually the calculus of random variables is employed as the "Law of Propagation of Error". This law involves assumptions which should be critically examined, not blindly accepted.

3.2 Internal evidence

The first category that involves induction only relies on evidence internal to the digital file. Some internal checks would cover closure of polygons or use of undefined attribute codes. Nearly all of these procedures rely on some form of independent source of information. In simple cases, the independent source is an external list of legal codes. In more complex cases, a data structure might provide dual independent coding so that errors can be isolated. In hardware design and signal processing error detecting and correcting codes are standard, but those applications are largely one-dimensional.

When refering to internal checks, the precise nature of the algorithm should be explained. It is not enough to say that the file is 95% (or even 100%) clean without specifying how this is counted. For some components, such as the logical integrity of the data structure, testing cannot be based on any more rigorous form of evidence.

3.3 Comparison to source

In the current state of cartography, digital data bases are usually developed from existing maps. In any event, there is some form of material used as source documents. One class of test procedures uses these source materials to verify the quality of the digital product.

If the source material is a line drawing, a line plot could be generated from the digital data base at the same scale. Graphic inspection could identify the grossest blunders, but more sensitive tests may be needed. Redigitizing of a sample of lines could provide evidence of repeatability in recording spatial components. Similarly, if the source material requires interpretation, an independent interpretation could establish attribute consistency.

3.4 Independent evidence

Of course, source material tests cannot determine errors in that source material. The most rigorous form of testing compares a digital product to information derived from an independent source of higher quality.

One issue for this type of testing is expense. Independent evidence of higher quality should cost more than the normal sources used, or it would be used instead. In some historical or fast-changing situations, an independent source may not be available. It should be entirely adequate to confine testing to a sample drawn form the whole data base. A part of a standard should be an appropriate sampling procedure as a guide.

Testing against independent evidence brings in the element of "ground truth" as final arbiter of quality. In studies of attributes, the classification or other measures should be verified by detailed inpsection, but there is a risk that a ground survey will make finer spatial distinctions than intended in the more exhaustive map. Standards should ensure reproducibility in this matter.

Similarly, postitional data can be verified by a survey or air triangulation. For points (bench marks, section corners) or objects such as property lines formed from such points, this procedure will be unambiguous. More complex procedures will be needed for more complex features such as ridge lines, soil boundaries, et cetera. With such features the choice of points is arbitrary, and a ground survey might discover unintended detail.

3.5 Form of standard tests

The standard for data quality will largely consist of procedures, organized in this ascending scheme of rigor, and addressing the components mentioned in section 2. The tests will be laid out in generic form - what information to collect and suggested statistical considerations. It would be counterproductive to specify exact procedures in exact operational detail, let alone thresholds of results required.

4. Reporting test results

A digital data product's quality report should continue beyond the lineage statement to describe tests performed and their results. The level of testing will be optional, but there should be pressure from informed users to maintain an active testing program. Quantitative measures obtained from these tests provide some of the clearest information on fitness for use.

A suggested checklist or reporting form will be prepared to communicate different kinds of test results.

5. Specificity

While the general form of quality testing has a clear logic, there are some additional issues involved. In the simplest situation, quality of position, attributes and structure are uniform. However, this ideal is not always applicable. For example, geodetic control is usually far from uniform; discriminating some attributes is much less precise than other cases. In addition, a data base will be updated, corrected and manipulated in bits and pieces. Quality characteristics will not apply uniformly.

At one extreme, it might be useful to call for accuracy measures attached individually to each data item. This would lead to a doubling of file bulk and undoubted resistance in adopting such a standard. Recognizing that this extreme is realistic for a few quality-concerned applications, more moderate solutions are needed. Methods must be devised to ascribe quality characteristics to units intermediate between whole data bases and individual records. A digital version of the "reliability diagram" present on some map series would be useful.

6. Data control

The quality of a digital data base is an ongoing process. It is possible to correct virtually any error and to improve accuracy through improved control or attribute reassessment. To ensure that this process actually occurs, a producer should specify some mechanism to handle feedback.

At the same time, the producer should make claims of copyright, ownership or security classification as required.

As the distribution of data changes from tapes to communication networks, data basses will become more dynamic. It may be possible to allow updating of a data base to become decentralized. For instance, the county surveyor will know as section corners are remonumented and given accurate coordinates. The state highway commision will know when its highways are relocated. These users of digital data should pass their new information on to other users. Any such diversity of users will introduce substantial difficulty in maintaining quality control and standards.

7. Gaps in knowledge

Data quality involves many imperfectly understood components. The assessment of classification accuracy is a fledgling field. Some methods have been advanced, but there is still no consensus. Though substantial studies cover some portions of positional accuracy, there are even gaps in this topic, particularly concerning complicated features. In spite of the vast number of digitizers, and the sharp competiton between alternative technologies, There are few studies of variability introduced by these different technologies under typical operational settings. These gaps in knowledge would preclude a rigid threshold type of standard, because the quality of current production is very hard to know in any detail. By contrast, the "Truth in labelling" approach means that the standard can adjust to realistic circumstances.

(p. 33 follows

Report of Working Group III - Cartographic Features National Committee for Digital Catographic Data Standards

by
Warren Schmidt
Cartographic Consultant
Rand McNally & Company
11200 Lagoon Lane
Reston, Virginia 22091

ABSTRACT

The Working Group met three times since its inception in August 1982. Eight members, representing government, academe, and private industry, are now contributing to the study of cartographic feature classification. The issues of scale independence, data organization, form of feature definition, and basic cartographic data were addressed and the Canadian Standards were examined. Six goals for features in a national standard were recommended: independence from symbolization or scale; universality; logical structure; single classes with multiple attributes; explicit definition; and derivation from basic feature sets. Future plans call for a prototype, assembly of feature sets and glossaries, and cost estimates for feature standards.

INTRODUCTION

The purpose of the Working Group is to study cartographic feature classification systems and to specify a model that is consistent and comprehensive. This is complicated by many incompatible existing systems and the diversity of applications. To help sort out the confusion, the Working Group was directed to complete three tasks and address four issues. The specific tasks were:

1. Examine Federal, State, and local mapping specifications agreements.
2. Examine Canadian Standards for Exchange of Digital Cartographic Data.
3. List the implications of dealing with "real world" features - time and cost.

The issues to be addressed were:

1. Scale Independence - is it possible--if not, where are the breaks?
2. Organization - hierarchal or something else? What depth and level of detail?
3. What form of feature definition? Should the feature code be separate from the attribute code?
4. Is there a set of basic cartographic data? What is it?

After the completion of the above, the next objective was to take one or two data categories through the envisioned system, e.g.: coastline definition in hydrography and land use/land cover.

MEMBERSHIP AND MEETINGS

The Working Group is now comprised of the following members:

Mrs. Mary Clawson, IIT Research Institute
Dr. Beth Driver, Technology Service Corporation
Mr. Richard Hogan, National Ocean Survey
Maj. Robert Jacober, Air Command and Staff College
Dr. Joel Morrison, U.S. Geological Survey
Prof. Robert Rugg, Virginia Commonwealth University
Mr. Warren Schmidt, Rand McNally & Company, Chairman
Mr. Frederick Tamm-Daniels, Tennessee Valley Authority

The first meeting was August 22, 1982, and subsequent meetings were held on February 4, 1983, and March 13, 1983. Additionally, considerable communication between members and other working groups took place. The four issues and the Canadian Standards were each discussed at length and agreements were reached within the Working Group. The examination of mapping specifications and agreements and the implications task were deferred for later consideration.

SCALE INDEPENDENCE

The paragraph in the Canadian Standards for Exchange of Digital Cartographic Data (Volume I, page 5) on scale integration was reviewed. This states that a scale-independent system is feasible, applicable to all map scales, and would facilitate exchanges between users. The only disadvantage cited was the length of the eventual classification system. The idea of a single universal classification system was acceptable to the Working Group but the linking with scale was found to be irrelevant. Feature classification is an attempt to describe the real world. But in the real world features are independent of their cartographic representation and graphic scale. A building is still a building, no matter what symbol or scale is employed. Therefore, the basis for cartographic features would be a universal feature classification system, independent of representation and scale.

DATA ORGANIZATION

The original issue questioned whether feature organization should be hierarchal or relational and just how should the data be stored for efficient retrieval. Following lengthy discussion, it was decided that because data is not retrieved in the manner which it is stored, data organization in the data processing sense was not relevant to features. What is needed for features is a logical coding scheme not necessarily tied to any existing formal hierarchal or relational model. The design should be open-ended, flexible, and provide for the entry of features and associated attributes. The system, once created, will then be accommodated by the organization of the cartographic data base.

FEATURE DEFINITION

Should features be separated from attributes? What should be the depth of classification? On reviewing this issue, the Working Group came to the conclusion that each feature group should be a single class and explicitly defined. If more than one definition for a feature exists, such as that for "shoreline" in the Canadian

Standards, the difference should be captured in the attributes. Thematic features, e.g., aeronautical, geologic, land use, should be classified according to standards for the theme displayed. This will make for compatibility and encourage adoption of the eventual system. Attributes should be multiple and appropriately describe feature characteristics such as structure, composition, mensuration, and serviceability.

BASIC CARTOGRAPHIC FEATURES

Originally this issue dealt with basic cartographic data, but the Working Group substituted "features" because the meaning of the word "data" is too broad. Is there a basic set of cartographic features--a foundation for our future classification scheme? Implied is a uniform and universal series, a perfect data set that obviously doesn't exist. However, most maps were originally derived from topographic maps or hydrographic charts. Why not start with those two map types as sources for our basic cartographic features? The features shown on those series are well documented, have stood the test of time, and apply to most scales. This "basic" information, however, would only be a start and additional features would be entered from sources at different scales and showing themes such as soils, climate, and population. In the case of thematic maps, the individual features should be classified according to the theme discipline if standards exist and are contemporary.

CANADIAN STANDARDS

The three volumes were individually examined and found to be a useful reference. The Scale Integration statement was examined in the Scale Independence Issue above. Of particular interest was the Dictionary of Topographic Terms. In this glossary, each feature was explicitly defined. Multiple definitions were noted and this was dealt with under the issue dealing with Form of Feature Definition.

GOALS FOR A NATIONAL STANDARD

From the discussion of the issues and the Canadian Standards, six goals for features in National Digital Cartographic Standards were identified. The Working Group agreed that in a national standard, features should be:

1. Independent of cartographic representation and scale.
2. Universal in nature.
3. Logically structured.
4. Consist of a single class with multiple attributes.
5. Explicitly defined.
6. Derived from basic topographic and hydrographic feature sets.

FUTURE PLANS

The Working Group plans over the long term to accomplish the following:

1. To prepare and test a classification schema that meets the identified goals.
2. To collect topographic and hydrographic feature sets and glossaries.
3. To estimate the time and costs involved in assembling and preparing feature sets and glossaries.

Toward those goals, Beth Driver will prepare a proposed classification scheme and feature lists and glossaries will be obtained from the USGS by Warren Schmidt, NOS by Richard Hogan, and at large by Mary Clawson. The local members of the Working Group will meet in May to review progress. Ways of testing the feature scheme and the potential of student help will be discussed at that meeting.

SUMMARY

A representative Working Group has been staffed and is now active. The assigned issues of scale independence, data organization, form of feature definition and basic cartographic data were addressed and the Canadian Standards were examined. From these discussions came recommendations that features in a national standard should be independent of symbolization and scale, universal, logically structured, have single classes with mutliple attributes, defined, and derived from basic feature sets. Future plans call for a protytype classification scheme, assembly of feature sets and glossaries, and estimation of time and costs to prepare a feature standard.

DIGITAL CARTOGRAPHIC DATA STANDARDS: ISSUES IN TERMS AND DEFINITIONS

A Progress report

Working Group IV, NCDCDS

Prepared by
Dean T. Edson

Work Group IV membership

Mr. Dean T. Edson, Co-chairman
E-Quad Associates, Placerville, CA.

Prof. Harold Moellering, Co-chairman
Ohio State University, Columbus, OH.

Mr. Erich Frey, Vice Chairman
National Ocean Survey, Rockville, MD.

Mr. Frank Beck
U.S. Geological Survey, Reston, VA.

Prof. Mark Monmonier
Syracuse University, Syracuse, NY.

Dr. Carl Reed III
Autometric Inc., Fort Collins, CO.

Introduction

Almost every major work on the subject of digital cartographic data uses, has included a glossary section to provide a basis for communication and understanding. Such glossaries appear in the literature as far back as the mid 1960's and have reflected, for the most part, the steady progress in, but not always a clear understanding of, digital cartographic system technology. And the temporal nature of such glossaries causes most of the terms and definitions to become obsolete almost before the principle being focused upon is understood.

One such work was the 1980 edition of the ICA Glossary of Terms used in Computer Assisted Cartography [1] which was published by ACSM. Here, 750 terms were identified and defined by an international team. In just three short years, this glossary has lost its state-of-the-art aura and slipped to what one might consider a basic primer of terms. Since that time a whole new vocabulary has come into existance and we find system dependent terms popping up in almost every discussion on the subject of numerical cartography. One major problem associated with the creation of terms and definitions is characterized when new terms are logically constructed of well known term components which leads to a perceived definition that will vary between author and reader to the extent that concepts are not precisely conveyed.

We can therefore conclude from past experience that standard (uniformlly perceived) terms and definitions are an important component if we want to discuss issues and transfer knowledge related to digital cartographic data.

It is clear that the efforts of the Standards Committee will surface terms which have not been previously defined, or if they have, not in a way which is universally acceptable.

This brings us to the goal of systematically collecting and defining new terms which result in the development of standards for Digital Cartography. The primary tasks thus become:

1. Defining the outer limits of the term Digital Cartographic Data so that we won't wander off into the outer limits of the universe in our quest for standards.
2. Compile a list of new terms and perferred definitions.
3. Testing and validating existing terms and definitions which currently appear in the literature.

Some of the background issues which we will have to keep in mind as we pursue our goal are as follows:

Terms and definition requirement
We need to keep the concept of standardization in the background and concentrate on obtaining a clear definition concensus. When we identify agreement, the classification as a standard will naturally follow. If we try to force a standard we know what the profession is likely to tell us to do with it. We can anticipate that almost every proposed new term and, especially the definition, will create controversy. Converging on a concensus may be difficult but must remain a targeted requirement.

Geographic References
Because of efforts by other groups who deal with types of spatial data, there has developed a wide-spread misunderstanding of terms used in defining geographic reference. As an example, we have identified works outside the cartographic profession which use the following terms almost interchangably:

- Geographic coordinates
- Geodetic coordinates
- World coordinates
- Earth coordiantes

-Map projections and local coordinates

I suspect we will serve our cause more effectively if we take the time to go back into what might be referred to a "skeleton closet" of terms and make sure we have constructed a solid foundation upon which to build.

Multiple terms with a single definition or visa versa

In a recent presentation, Eric Anderson (USGS) developed a matrix of terms and sources which help focus on this issue. As an example, he lists some of the terms found in recent literature which define "a line". His list includes:

Line
Arc
Chain
Segment
Edge
Face
Link

Some of these terms are system inspired and may also have a deeper or more complex meaning than "line" so we will have to deal with this problem carefully and insure that term interrelationships are fully developed and explained.

Buzz words and shop jargon

These terms abound in the land of digital cartographic systems. The intended purpose might include but is certainly not limited to: explaining, confusing, protecting copyrights, or a variety of other possible motives. But we have to keep in mind that, whatever the reason, in some cases yesterday's buzz words may become tomorrow's standard term. One such example of jargon turned into a standard term may be "cartographic spaghetti". Again, if a term is clearly understood and perferred we may be well advised to call it a "home run".

Credibility

There is a deep rooted feeling in the private sector that anything emanating from Federally funded, coordinated or inspired effort is somehow bad and thus will be met with a healthy skepticism. The decision making process must and will, therefore, be so open and clear that anyone will be able to follow the logic of our decision process and will understand, and, hopefully, accept the results.

Gaps in our Knowledge

We must be alert to "Gray Area" in our knowledge and avoid rushing in with new so called standard terms as we explore the unknown. In many cases, the test of time will be needed to determine if logic, function and performance will have prevailed.

With these goals and issues in mind, the Terms Work Group looks forward to the exciting and perhaps not impossible task which lies ahead.

References

1) Edson, D.T. / Denègre, J., A Glossary of Technical Terms in Computer Assisted Cartography, American Congress on Surveying and Mapping, Falls Church, VA. 1980.

General Comments on the Work of the National Committee for Digital Cartographic Data Standards, Cycle 1

NAME: ______________________ WORK PHONE: ______________________

ADDRESS: ______________________ EVENING PHONE: ______________________

____

____

____

Your Comments Please:

SIGNED: ______________________

DATE: ______________________

Cut Here