EDITOR'S INTRODUCTION

These seven volumes summarize the proceedings of the First International Symposium on Topological Data Structures for Geographic Information Systems, held at Endicott House, Dedham, Massachusetts, U.S.A. between October 17 and 21, 1977. At these meetings more than seventy participants from seven nations convened for over thirty hours to discuss fifty-seven papers; this was the first time that so many active researchers had convened to exchange views and report on technical developments concerning data structures for spatial information. Although many of the participants were previously acquainted with each other and their work, having collaborated, corresponded, and attended other meetings together, no previous convocation had permitted such detailed examination of topics of mutual concern. Many of the participants had had no prior opportunity to present their work to other enthusiasts of the youthful but burgeoning field of geographic information systems.

What are "topological data structures," what is meant by "geographic information systems," and what connection do these concepts have? As witnessed by the highly variable use of terms and concepts within these proceedings, no single general definition can encompass the wealth of ingenuity with which we are presented. Webster's 3rd offers this definition of topology: "a branch of mathematics that investigates the properties of a geometric configuration that are unaltered if the configuration is subjected to any one-to-one continuous transformation." Mathematicians may take issue with this, but whether accurate or not, this definition misses the point of why topology is an important property of geographic information. The issue is not what occurs when space is deformed, but the ways in which objects in space relate to one another, their spatial properties. Even in the discrete world of digital data, space is usually represented as a continuum in which places have connections and relationships of various sorts. Geographic relationships may be expressed geometrically ("Boston is northeast from New York" or "Boston and New York are 188 air miles apart"), in which case some sort of measurement (in some explicit metric) is required to quantify the relationship. Geographic relationships may also be expressed topologically, with no measurement required ("35 flights a day link Boston to New York" or "the next train from New York to Boston stops at 11 stations"), yet still quantitatively.

Whether one considers geographic phenomena as points, lines, areas or volumes, topological principles can be applied to describe their relationships. City blocks, streets, and street intersections have mutual topological relationships which can be logically and quantitatively described, and which in turn can locate urban phenomena independently of any coordinate system relating to the earth's surface. River tributaries connect in ways which permit their topological ordering and calculation of characterizing statistics such as bifurcation ratios. Landforms, which have a topography, also have a topology in which all critical features (peaks, pits, passes, ridges, courses, hills, and dales) are tied together, and may be represented as a graph as well as a surface. Topology thus expresses properties as useful as metric length, area, flow, and density, and perhaps more intrinsic to spatial systems. It may be flatly claimed that no characterization of a region can be considered adequate without reference to topological notions of dimensionality, adjacency and connectedness.

Computer cartography has provided much of the impetus for the development of topological data structures, although only recently have the geographic data bases used for plotting maps themselves incorporated topological principles. Early cartographic base files tended to describe all objects, whether points, lines or areas, simply and independently. This limited their use to reproducing maps, and hindered error correction. When boundaries of zones -- or "polygons" -- are described as independent lists of vertices, each point on a shared boundary should have at least one duplicate elsewhere in the file. If such a point is added or deleted, its neighbor's description ought to be similarly modified. But if each polygon is separately described, there is no straightforward way to accomplish this. With the explicit encoding of topological information (which need not add bulk to the file, and may permit it to be reduced in size) adjacent and connecting objects name one another, making it obvious which objects are affected when a given object is modified. Topological encoding thus yields immediate practical benefits to computer cartography, as well as permitting many different objects (and their relationships) to be described systematically in one geographic file. Both hierarchical (e.g., U.S. counties and their aggregates such as states) relationships and overlaid ones (e.g., soil and land use zones) benefit commensurately when described in topological language. Applications of such data can then extend beyond drafting maps to include selective retrieval, aggregation, logical comparison, line generalization, address matching, and flow modelling. All of these capabilities are described as operational in various implementations by Symposium authors.

The papers themselves touch many issues, centered naturally on the application of topological principles to computer manipulation of geographic data. But as even a casual reading of their titles will indicate, other topics both broad and narrow found their way into the agenda. Thus we find discussions of program user command languages, errors in isarithmic mapping, pattern grammers for maps, and technology transfer in these proceedings in addition to the expected mix of papers concerning DIME file applications, polygon overlay, interactive cartography and algorithmic complexity. Likewise, many contributions dealt with particulars of data structures, but a substantial number of them made little or no reference to how information is organized by software. Those which did, dealt diversely with various dimensionalities of data -- points, lines, areas, surfaces, and on into N arbitrary dimensions. Sometimes topology was explicitly expressed in a data structure, and sometimes existed by implication only. Such distinctions were pounced upon by some participants and discounted by others; opinion varied greatly concerning the necessary and the sufficient elements of data structures, depending on whether one was primarily concerned with data acquisition, processing efficiency or ultimate application, or whether one was a theoretician, programmer, planner, or manager. Some papers describe existing software or data bases, others specify performance criteria for geographic information systems. Specific analytic and cartographic applications are reported in some cases, and in others only general capabilities are outlined.

The idea for the Symposium first came to light in the Spring of 1976. An increasing amount of correspondence, published papers, government documents and informal reports concerning geographic data structures and their applications had been accumulating in the Harvard Laboratory for Computer Graphics and Spatial Analysis (LCG). Spatial data processing, still a new

sub-discipline connecting research in computer science and geography to policy
and planning-oriented decision making, was clearly leaving the sheltered
environments of research labs and moving into field applications. A number of
federal and state agencies had developed prototype and production geoprocessing
and mapping systems which were computer based, and levels of interest were
rising in all sectors, both in North America and abroad. Such interest was
manifested at conferences such as the Auto-Carto I and II meetings in Reston,
Virginia, by the growth of the SORSA (Segment-Oriented Referencing Systems
Association) organization, which brought together practitioners in North
America and Europe, and by mounting requests for technical documents and
software known to or generated by the LCG.

Much of this material comes under the category of "fugitive literature";
mimeographed and privately circulated, hidden in obscure government documents,
or occasionally published in journals principally devoted to other
specialties. No recognized, refereed journal concerning geographic
information systems or their theoretical foundations was (or is) available as
a convenient channel for discussion. In consequence, a lot of effort was
expended tracking ideas back to sources, and much research involved
unnecessary duplication of work. Ominous grumblings would erupt from time to
time from researchers who felt their contributions had been ignored, or even
stolen. Some called for a new professional society, others for a new journal
or both; but until a community of practitioners was better identified these
suggestions seemed premature and difficult to implement. Although the field
of research workers is larger than is generally realized, and quickly
expanding, it is still small enough for individual work to receive attention
by a good proportion of interested colleagues, given a suitable forum.

Allan Schmidt must receive the major credit, not only for organizing the
Endicott House meeting, but for first perceiving that the time had arrived
to come together to communicate research results, explore unanswered questions,
systematize modes of discourse, and in general to assess the state-of-the-art
and its prospects. His judgement and efforts, intensively assisted by Nick
Chrisman, Denis White and others at the LCG, were significantly responsible
for the high level of discussions during the Symposium sessions. Ten to
fifteen papers were discussed each day, using a somewhat novel approach to
this task. Rather than having authors summarize their own papers, volunteer
discussants were appointed to do so, and encouraged to set forth particular
issues for general discussion (based on the content of each paper and the
theme of the session). The summaries were generally brief, since
participants had submitted papers well in advance and had received workbooks
containing all the papers one or two months prior to the Symposium. Thus
armed and forewarned, the participants' grasp of the material was strengthened,
and discussions avoided being dominated by tedious particulars or vacant
generalities. Having a moderator to police each discussion helped insure that
everyone who had something to say was recognized, and discouraged shouting
matches. Controversy was inevitable, given a mixture of research scientists,
educators, research sponsors, and applications people. This proved healthy
and stimulating, expanding the worms-eye-view of the "boffins" described by
Brian Berry in his welcoming address, and advising everyone on the limits of
the possible. It is hoped that some of this flavor can be transmitted to a
much wider audience through publication of these proceedings.

In the hope that structuring the 1800 or more pages of contributions to the Symposium can make this material more accessible, the proceedings are being published in a seven-volume series. This ordering of papers differs somewhat from that used during the Symposium, which was divided into eighteen sessions (see reprinted program, pp. 11 - 16). Volume I contains none of the original papers; it is intended as a summary overview and reader's guide to the content of the Symposium. In it are found day-by-day summaries of discussions, along with abstracts for all papers. The summaries are based on notes furnished by the Symposium rapporteurs (participants who volunteered to monitor discussions, cloistering themselves at the close of each afternoon with their notes and the taped record of discussions to produce daily abstracts of the proceedings), but contain little of this material verbatim. In rewriting the summaries, the Editor has reviewed most of the taped transcriptions, and relied on his own recollections in an attempt to document recurrent themes and controversies and to enforce some modicum of stylistic consistency. The task of distilling over thirty hours of frequently abstruse dialogues into a few score pages was both difficult and rewarding: if the result of this process is more of a murky residue rather than a clear essence, the Editor -- not the rapporteurs -- must accept responsibility.

Volume I begins as the Symposium itself did, with a keynote address by Dr. Brian Berry, Director of the Laboratory for Computer Graphics and Spatial Analysis. In welcoming all participants, Dr. Berry set forth concerns about the applicability of geographic information systems to local planning efforts, concerns which few papers had directly addressed but which indeed threaded through the ensuing discussions. Toward the end of Volume I are found two contributions -- by Nick Chrisman and Jim Corbett -- not discussed at the Symposium but offering overviews of its content, appropriate for publication in this summary of proceedings. Chrisman's was written on the eve of the Symposium and distributed as part of the working papers. It attempts to synthesize views of various authors concerning topological data structures into some manageable paradigms without actually becoming a critique. Corbett's brief memo is more restricted in content, and addressed what he feels is the essence of the topological approach to geographic phenomena. In it he explores the hypergraph model of Francois Bouillé, relating it to research undertaken by himself and others at the U.S. Census Bureau. Though not a general overview, this precis has been included in order to amplify understanding of Bouillé's contribution -- a thoughtfully conceived, highly general and powerful approach to topologically-based modelling, different in both degree and kind from other reporter approaches to topological data structures. Due to the difficulties of presenting such abstract concepts in a foreign language, comprehension of Bouillé's work was somewhat hindered; Corbett, who unfortunately has no contribution of his own in these proceedings, has rendered a service in providing this commentary. Also found in Volume I are participants' contributions to the "Future Directions" sessions. These discussions wound up the Symposium with exchanges of ideas concerning the importance of and influences upon geographic information systems as tools for social progress. Brief statements prepared by participants, mainly during the course of the Symposium, were used as context for these sessions, which dealt variously with evolving technology, emerging social and educational issues, avenues for cooperation and transfer of systems and their components, and the elements of the "real world" which information systems purport to describe. All written contributions to these

sessions as furnished by participants have been reprinted in Volume I
following the summaries of the Friday sessions.

Note that all discussions are summarized chronologically (although in the
case of Tuesday's sessions the commentary reorders some sessions to reflect
similarities among morning and afternoon discussions), and that paper
abstracts precede the day's summary in which they were initially discussed
(some papers were discussed more than once). Each of the five sets of
abstracts is ordered alphabetically by author, and an overall Index to
Abstracts appears at the end of Volume I. Finally, there are six checklist-
type indices to the papers themselves, for Volumes II-VII of the proceedings.
This format was chosen to make the content of the papers accessible to
readers on a comparative basis. Though it does not purport to be an
exhaustive enumeration of all issues addressed in all papers, this index
scheme highlights similarities and differences among the papers according to
some major recurring themes of the Symposium. Clearly this represents a
rather incomplete and somewhat subjective approach to the subject matter of
the Symposium as well as to the content of the papers themselves. Its
purpose (as is that of Volume I in general) is to provide a springboard from
which readers may launch their own inquiries, and no guarantee of
comprehensiveness or authoritativeness is expressed or implied. The richness
of the contributions to the Symposium clearly defies any attempt to summarize
them in a list of keywords, no matter how extensive the list may be.

Volumes II through VII of the Proceedings are devoted to reprinting the
working papers themselves. They are structured by topic and subtopic, into
three pairs of two columns each. Certain papers obligingly restricted their
content to specific issues, and were therefore readily classified, while
others touched on a variety of concerns, or were sufficiently unique that
assigning them to any particular category is problematical. The difficulty
of classifying any particular paper can in some cases be measured by the
categories it spans in the checklist indexes found at the end of each volume.

Geographic Information Systems: Surveys, Overviews and Criteria is the title
of Volume II. This set of papers is perhaps the best starting point for
readers desiring general appraisals of the state-of-the-art and the salient
issues concerning the use of information systems with geographic data. The
eleven papers in Volume II variously describe operational or prototypical
information systems, certain data resources, general requirements for systems,
and problems involved in the transfer of such technology and data. Some of
these inquiries lead to philosophical questions concerning the proper
application of spatial analysis techniques. Such broader questions are dealt
with directly in some of the papers found in Volume VII, described below.

Moving from general capabilities of geographic information systems to specific
problems solved by them, we come to Volume III, Geographic Information Systems:
Cartographic and Analytic Applications. This may be regarded as a companion
to Volume II, as its contributors all describe specific systems, but in actual
problem-solving contexts. Several of these works deal solely with cartographic
applications, but the majority describe uses involving spatial analysis or
land inventory. Most are quite explicit in their descriptions of data
structures employed, while the remainder prefer to stress the kinds of
analyses which topological data bases enable their systems to perform.

Volume IV, <u>Data Structures:  Hierarchical and Overlaid</u>, also contains material relevant to the two prior volumes, as in most cases operational systems handling respectable quantities of data are described.  In these papers, however, the focal point tends to be how data are organized for coding, storage retrieval and processing.  Not all the data structures described are topological, some concern only sets of points, but most commonly what is of concern are areal coverages, or polygon sets.  The distinction made in the title applies to many of these papers:  either the data structure is optimized to deal with polygon hierarchies (nested zones, such as counties and states), or with overlaid polygons (different coverages overlapping in arbitrary ways).  Hierarchical data are more tractable, and some clever approaches to spatial searching are described which take advantage of the properties of hierarchies.  Overlaid data is described as being stored using grid cells, topologically, or using some mixture of the two; several hybrid systems are explained.  Readers with particular interest in spatial overlay problems are also directed to Volume VI (on Algorithms) where the computational issues of overlay are explored by several authors.

Data structures which describe surfaces are of principal concern in Volume V, <u>Data Structures:  Surficial and Multi-dimensional</u>.  Most of these papers describe systems for handling physical terrain data, in some cases elucidating the connections between topology and topography.  The topology of surfaces, first applied to geography about twelve years ago, has nourished new approaches to digital terrain modelling, and the fruits of this infusion of theory are evident in the papers found in Volume V.  Most of these papers concern terrain modelling, and of these most schemes use topological data structures (triangulations), but gridded data structures are also discussed.  Two papers deal with the analysis of surfaces, and one (Bouillé, mentioned above) addresses not surfaces <u>per se</u>, but describes a data structure and theoretical framework for modelling arbitrarily complex phenomena in many dimensions.

<u>Spatial Algorithms:  Efficiency in Theory and Practice</u> brings together a group of papers which treat a diversity of problems, but all of which express direct concern for the efficiency of algorithms from both theoretical and empirical perspectives.  As demonstrated by comments elicited during the Symposium's algorithm sessions, there as yet exists little agreement concerning the proper measurement of how well a program performs complicated tasks.  Thus some papers discuss efficiency in terms of "asymptotic complexity," a theoretical order-of-magnitude approach, while others present statistics arising from actual processing runs.  Regardless of how it is approached, the quest for fast algorithms is important to the creation of effective information systems.  Such studies can indicate the expected limits to the amount of data which can be pushed through a system, and theoretical work can identify whole classes of problems which are insoluble in principle, or which must be reformulated in order to be solved.

We conclude these proceedings in Volume VII, <u>Spatial Semantics:  Understanding and Interacting with Map Data</u>, with perhaps the most diverse group of contributions to the Symposium.  Ranging from descriptions of command languages for program control to epistemologies of space, the spectrum of concerns also includes DIME file editing, digitizing protocols, "linguistic" analysis of maps, the information content of map data and a "science fiction" geographic display system being developed by the U.S. Department of Defense.

Despite such diversity, all these papers have at least one common theme; how users of spatial data conceptualize and manipulate their data. Whether the focus is the nitty-gritty of batch processing in a federal agency or an attempt to understand how the human mind processes spatial information, all these papers take the human element into account in discussing connections between the "real world" and various abstractions of it. More than any other group, the papers in Volume VII illustrate that geographic information systems, regardless of their sophistication, are mirrors not of the world, but of our understanding of the world. The Symposium was undertaken, and these proceedings are offered, in the hope that this distinction will continually be clarified, and eventually be made to disappear.