When Data Worlds Collide
Refereed

3 Object-Oriented Modeling for GIS
    Max J. Egenhofer and Andrew U. Frank

20 Benefit-Cost Principles for Land Information Systems
    Gregory L. Poe, Richard C. Bishop and Jeffrey A. Cochrane

32 Assessment of TIGER/Line Files for Rural Counties: The Case of Hays County, Texas
    Ryan Rudnicki

47 Integrating Dynamic Modeling and Geographic Information Systems
    N.W.J. Hazelton, F.J. Leahy and I.P. Williamson

59 Systematic Development of Education and Training Programs: A Key to Successful GIS Development
    Charles R. Blinn, Lloyd P. Queen, Lorentz R. Hegstad, and Daniel J. Fitzpatrick

Features

69 GIS-ing About in Public: An Anatomy of a Public Disgrace
    John Keene

71 Why We Couldn't Get the Data We Wanted
    William J. Craig

79 Alberta's LRIS in Business—Part 1: Preparations
    Patrick J. Merrick and Laurel J. McKay

85 Horwood's Short Laws
    Stephen Kinzy

Feature Map

91 Applications for Digital Terrain Modeling
    Robert S. Walker
Reviews

Books

88  An Introduction to Urban Geographic Information Systems  
    William Huxhold (Reviewed by Kenneth J. Dueker)

Software

94  Dynamic Segmentation Revisited: A Milepoint Linear Data Model  
    Kenneth J. Dueker and Ric Vrana

106  Taking the TIGER by the Tail: A Software and Data Review  
    Alan A. Lew and Richard E. Krisman

118  In Depth: Atlas Pro for the Macintosh  
    William Casey

129  Index

134  Guidelines for Manuscript Submission

Cover: Sharing data has many benefits, yet it's often difficult to do. Our cover illustrates one problem inherent in data sharing: the two datasets have different geographic bases, so they can never connect. Will Craig has studied the real-world problems of sharing data. His findings are illustrated in 29 cartoons and line drawings in an article starting on p. 71. Using pictures and humor, Craig illustrates different aspects of a common problem—that people cut corners when creating databases because they're focused on their own needs. The intent of Craig's article is to dramatize the importance of high-quality data and to facilitate more sharing in the future.
EDITORIAL BOARD

Robert Aangeenbrug  
University of South Florida

David Arbit  
City of Austin

Kate Beard  
University of Maine

Richard Brail  
Rutgers University

Charles Drinnan  
Synercom Technology, Inc.

Earl Epstein  
Ohio State University

Lewis Hopkins  
University of Illinois

William Huxhold  
University of Wisconsin, Milwaukee

Ken Jones  
GIS Consultant, Oslo, Norway

Charles Kindleberger  
City of St. Louis

Richard Klosterman  
University of Akron

Kenneth Kraemer  
University of California-Irvine

Robert LaMacchia  
Bureau of the Census

John Landis  
University of California-Berkeley

Robert Lima  
Boshe Institute

John McLaughlin  
University of New Brunswick

Gilbert Mitchell  
National Geodetic Survey

Joel Morrison  
National Mapping Division, USGS

David Phillips  
University of Virginia

Carl Reed  
Genasys, Inc.

Heinrich Reimermann  
Hochschule Für Verwaltungswissenschaften

Vincent Robinson  
University of Toronto

Eugene Roe  
DCA Softdesk

Mark Salling  
Cleveland State University

Allan Schmidt  
Schmidt Associates

K. Stuart Shea  
TASC

Larry Sugarbaker  
Washington Department of Natural Resources

Barry Wellar  
University of Ottawa

Peter Zwart  
University of Tasmania

Paul Zwick  
University of Florida

MANUSCRIPT SUBMISSIONS: URISA Journal is dedicated to sharing knowledge about information systems among managers, users, developers and educators, so that improved systems can be developed and used more effectively and equitably at all levels of government. Manuscripts, correspondence or questions on editorial matters should be addressed to the following section editors: Referee: Ken Dueker, Portland State University, P.O. Box 751, Portland, OR 97207; (503) 464-4042. Features: Gilbert Castle, 44 Montgomery St. Suite 500, San Francisco, CA 94104; 415-955-0531. Reviews: Rebecca Somers, PRC, Inc., 1505 Planning Research Drive, TM615, McLean, VA 22102; (703) 883-8422. Maps: William Craig, Center for Urban and Regional Affairs, 330 Humphrey Center, Minneapolis, MN 55455; (612) 625-1551. Section editors will provide Guidelines for Manuscript Submission upon request.
EDITORIAL INTENT

The refereed section of URISA Journal strives to share new knowledge in the technical, social, economic, and institutional subject areas that support information systems technology. It is the intent that this section of the journal contain papers that are representative of URISA's membership and the broader information systems community. We encourage the participation of system designers, implementors and users as well as the educational and research community.

We hope that the refereed section will provide reliable information and new insights resulting from experience, research and scholarship. We also hope that this section will link academia, industry and the user community through the sharing of critical investigations and organized knowledge. To this end, we are seeking three forms of work: (1) Reports of current research and development pertinent to the overall information systems community; (2) systematic literature reviews of research for the research and development community; and (3) systematic reviews of applications which explain successful systems and procedures to the overall information systems community.

The refereed process consists of a "blind review." After receiving a manuscript from an author, we send it out for review to three or more persons who have been identified as being knowledgeable in the topical area. The name and affiliation of the author are removed from the paper so the reviewers can give it an impartial review; likewise, the names of the reviewers are not revealed to the author. We ask the reviewers to respond to the following: (1) Is the thesis or purpose stated early and clearly; (2) is the significance of the paper stated explicitly; (3) is the thesis argued persuasively; (4) is the writing clear, concise, straightforward, interesting, and in the active voice, where possible; (5) is the paper tied in appropriate ways to relevant literature; (6) is the paper illustrated appropriately; (7) are the methods sound and appropriate to the paper; (8) are the methods explained clearly; and (9) is the paper interesting to many different types of URISA Journal readers?

If the manuscript is accepted and does not need revising, it is sent to the managing editor for comments and final editing. If the manuscript needs revision, assistance is provided by the editors. If the manuscript is not accepted, it is returned with an explanation by the editors.

(For complete guidelines regarding the preparation of manuscripts and illustrations, see p. 134.)

Kenneth J. Dueker

In This Issue: Something for everyone. The refereed papers in this issue are largely unrelated, sharing neither common theme nor approach. Yet each contributes knowledge about development and use of information systems.

Egenhofer and Frank provide a reasoned review of the potential and problems of moving to greater object orientation in GIS. Their work serves as an organizing framework for readers familiar with the topic, and its clarity enables those unfamiliar with object-oriented approaches an entry point to that literature.

Stark fills the rural void for TIGER bulletin, specifically identifying the problems in using the TIGER data for local applications. It is a tool for organizations that plan to use TIGER outside of urbanized areas.

Blinn, Queen, Hegstad and Fitzpatrick address an increasingly important issue GIS education and training. They employ a case study of development in the Minnesota DNR with emphasis on a learner needs assessment.

Poe, Bishop and Cochrane provide a rigorous discussion of benefit/cost principles for land information systems. They show how estimates can be biased in either direction. They employ a welfare economics framework and place the emphasis on estimating social benefits.

Hendron, Leahy and Williamson review requirements of dynamic modeling and call for developments in GIS to incorporate time and elevation. They refer to 4-D GIS as a "grand challenge," that will allow modeling and visualization of future scenarios.
Object-Oriented Modeling for GIS

Max J. Egenhofer and Andrew U. Frank

Max J. Egenhofer is an assistant professor in surveying engineering at the University of Maine and a researcher with the National Center for Geographic Information and Analysis. He holds a Ph.D. from the University of Maine and a Dipl. Ing. from Stuttgart University. His research interests are in spatial reasoning, the formalization of spatial relationships, spatial query languages, and the design of database management systems for GIS.

Andrew U. Frank is the chair of the Department of Geo-Information at the Technical University, Vienna, Austria, and a research professor in surveying engineering at the University of Maine. He received his Ph.D. from the Swiss Federal Institute of Technology, Zurich, Switzerland and a Dipl. Ing. from the same institute. His research is related to geographic information systems, particularly the handling of geometric data in computer systems, increasing the modeling potential of database management systems, and improving the human interface of a GIS, including visualization of geographic data and cartography.

Database management systems (DBMS) play a central role in a geographic information system (GIS), because they liberate the GIS designers and engineers from building and maintaining a substantial and complex part of a large software system (Frank 1988a). DBMSs have become an accepted tool for storing data in a form that is concurrently accessible for multiple users, prevents loss of data, and provides security control for access (Codd 1982). By using a database management system, all the tasks of low-level data management are removed from the programmer's and end-user's responsibility. From a programmer's perspective, data are described by their logical properties, not the physical structure in which they are stored. For instance, database users may retrieve the object "town of Orono" by providing the name "Orono" as a prop-

Abstract: The data model upon which most of today's commercial database management systems are based has shown to be insufficient for geographic information systems (GIS). The recently promoted object-oriented model provides some useful tools for data abstraction and data structuring, which augment the conventional tools and overcome some deficiencies inherent to the traditional relational model. In particular, the concepts of complex objects and pertinent operations are more powerful modeling methods than the currently popular structure of relational tables and relational algebra. This survey article presents the concepts of object-oriented modeling applied to geographic data and demonstrates their impact on future GISs.

Note: This work was partially funded by a grant from Intergraph Corporation and grants from Digital Equipment Corporation under TP-765488 and Sponsored Research Agreement No. 414. The support from NSF for the National Center for Geographic Information and Analysis under grant number SES 88-10917 is gratefully acknowledged.
property of an object of the class “town”; they need no knowledge about the location of the particular record in the file or details about the access method. The DBMS as a subsystem of a GIS can be replaced by another product of the same modeling power, provided the interfaces are compatible.

Computer scientists have studied the design and implementation of DBMSs for many years and several commercial systems are currently available. They are generally based on one of the classical data models—hierarchical (Tischritzis and Lochovsky 1977), network (CODASYL 1971), or relational (Codd 1982)—or their derivatives. Within the last few years, commercial relational database management systems have gained much popularity, and products such as Oracle and Ingres are widely used. They are built on the relational model, which organizes data as tables or relations (Codd 1970). Columns of the tables are called attributes and all values in an attribute are elements of a common domain that describes the set of all possible values. Rows are referred to as records, tuples, or relation elements (Ullman 1982).

While relational database management systems are suitable and successful for applications dealing with weakly structured data such as banking accounts and personal files, they fail when they are used for applications of data with a complex structure. GISs, which integrate data from various resources into a single, homogeneous system, need powerful and flexible data models to serve multiple tasks. These include:

- Sophisticated treatment of real-world geometry (Frank and Kuhn 1986; Greene and Yao 1986; Herring 1987; Egenhofer, Frank and Jackson 1989; Milenkovic 1989);
- Representation of the same data at different conceptual levels of resolution and detail (Bruegger 1989; van Oosterom 1990; Buttenfield and McMaster 1991);
- Management of history and versions of objects (Sernadas 1980; Snodgrass 1987; Langran 1989; Al-Taha and Barrera 1990); and
- Combinations of measurements of different resolution and accuracy (Davis 1986; Buyong, Kuhn and Frank 1991).

The GIS literature (Morehouse 1990; Frank and Mark 1991) often differentiates between systems that treat objects with distinct identities and raster systems, which store the distribution of properties in space (Tomlin 1990). We restrict our discussion here to GISs that handle objects with a distinct identity.

Since the relational data model does not match the natural concepts humans have about spatial data, users must artificially transform their mental models into a restrictive set of non-spatial concepts. For example, it imposes too many restrictions, such as normalization rules (Codd 1972), so that spatial objects must be artificially decomposed into smaller parts (Nyerges 1980; Frank 1988a). This implies a performance degradation of the databases that are based on this model, proportional to the size of the data collection, and impedes the use of traditional database management systems for geographic applications. This becomes particularly visible during interactive graphic sessions when large amounts of spatial and non-spatial data must be accessed to produce an understandable rendering (Egenhofer 1990) and users have to wait unreasonably long for the responses to their queries (Frank 1981). Though these deficiencies reveal performance problems, they are clearly based on conceptual issues rather than on hardware limitations. Problems that can be solved by exploiting additional or faster hardware will not be addressed here, since the requirements for these spatial database management systems are not hardware issues (Frank, Egenhofer and Kuhn 1991). For example, even with much faster access to storage devices, spatial indexing structures and large object buffers will be needed to provide adequate performance for queries, for which the results are to be presented as interactive drawings (Smith and Frank 1990). Faster and less expensive CPUs, larger hard disks and more computer memory are available, but they do not overcome the data structuring limitations in database technology applied to problems in spatial data handling. It is assumed that hardware technology will continue its rapid growth over the next decade, but substantial performance improvements of hard disks are not expected. Disk I/O is recognized as the major performance bottleneck in database management systems, and improvements by clustering and shadowing several disks are
only marginal remedies (Stonebraker et al. 1988).

Several branches in computer science (artificial intelligence, software engineering, database management systems, human–computer interaction) have recently promoted an object-oriented approach.

- Object-oriented data models have been developed to capture more semantics than the relational model (Brodie, Mylopoulos and Schmidt 1984; Peckham and Maryanski 1988).
- Object-oriented user interfaces make systems appear more natural and easier to use (Schmucker 1986).
- Object-oriented database management systems have been investigated to provide the corresponding features for storage and retrieval of complex objects (Zdonik and Maier 1990).
- Object-oriented software engineering techniques and programming languages have been developed to support the implementation of software systems that were designed following an object-oriented approach. They allow for immediate implementations of object concepts rather than simulating them with traditional programming languages (Stroustrup 1986; Meyer 1988).

The goal of this paper is to familiarize the reader with the power of object-oriented modeling for spatial data handling. The focus will be on conceptual achievements that will help to improve the modeling power of spatial information systems so that the often complex spatial phenomena may be expressed in terms closer to humans' thinking. Attaching the adjective object-oriented to a system just because it was implemented in an object-oriented programming language is misleading. Such a detail should not be visible to the user and is thus irrelevant (Herring 1992). Unlike a paper with a similar title (Worboys, Hearne and Maguire 1990), which stresses the structural aspects of a very specific object-oriented model (Abiteboul and Hull 1984), this paper attempts to present a synthesis of concepts common to different object-oriented models, and focuses on behavioral aspects.

This paper proceeds as follows: First, an overview of the need for formal models in GIS is given. Then an object-oriented data model is introduced that builds upon the four major abstraction methods of classification, generalization, association, and aggregation. The behavior of objects is explained in terms of the concepts of inheritance and propagation to model dependencies among objects in generalization and aggregation hierarchies. Finally, the object-oriented data model is compared with the traditional relational data model.

Modeling for GIS

Geographic information systems serve as repositories of observations humans make about spatially related objects and their properties. In order to concentrate on the issues necessary to solve a given task, humans build mental models of real objects and further simplify reality by using abstraction mechanisms until only the necessary components remain. Such mental models are either informally communicated by using natural language, or in a system based on a formal abstract model of reality.

Experientialists have observed that the base concepts humans use are established early in their life as abstractions from their bodily experience (Johnson 1987). This experience is the same for all humans, as they depend primarily on the physiology of the human body, and establishes a foundation on which to base communication and to avoid subjectivism. The meta model is the generic description of a situation (Ellis et al. 1993) (e.g., “a person owns a building” as opposed to the specific model “Doe Smith owns the building at 56 Park Street”). The abstraction mechanisms available in the data model determine the meta models and specific models which can be used, and thus the expressive power of a GIS (Figure 1). If the abstraction mechanisms are insufficient, the meta model of reality is inadequate and the mapping from the user concept of the object behavior onto the GIS model will be strenuous and difficult to understand. This makes the GIS hard to use.

Data Models and Abstraction Mechanisms

An important part of the model-building facilities is contributed by the data model used for the database management system in the GIS. A data model is a collection of (1) data structure types, (2) operators or inferencing rules, and (3) general integrity constraints (Codd 1981). It provides the tools, i.e., the lan-
language available, to describe the meta model of a database (Date 1986). Examples of data models are: the relational model (Codd 1970), the Entity-Relationship-Model (Chen 1976), first-order predicate calculus (Gallaire, Minker and Nicolas 1984), and object-oriented models (Manola and Dayal 1986; Bancilhon et al. 1988).

The data model of a database management system must provide for abstraction mechanisms that are also carried forward into the programming language and the software engineering environment; therefore, software engineering must provide the users with tools to apply the abstraction mechanisms immediately in a programming language. For example, if a model offers a mechanism to establish objects that can be composed of other objects, then the programming language used for the implementation must offer a construct to get all components that are part of another component. The lack of appropriate constructs in programming languages often leads to simulations that make software systems complex and difficult to maintain.

It is standard practice that an information system is designed using an Entity-Relationship-Model (Chen 1976), then implemented in an Algol-like programming language, e.g., Pascal or C, which was extended to access a relational database management system through an embedded query language such as Embedded SQL or Embedded Quel. Incompatibilities occur at each interface between two sets of tools due to the different models for information representation. This confusion of various data models results in what has been called impedance mismatch. Powerful features of one component must frequently be simulated in another. These simulations reduce efficiency and lead to discrepancies between the design and the implementation. The resulting product, based upon an incoherent design, consumes undue resources and is difficult to maintain.

For example, the calculation of the area of a 100-foot buffer along Route 1A has to be broken down into the following steps:

- Retrieving the line segments of Route 1A from the relational database with an SQL query into a “cursor”:

```
DECLARE road CURSOR FOR
SELECT start.x, start.y, end.x, and end.y
FROM road, edge, node
WHERE road.name = '1A'
and
road.lines = edge.id
and
edge.start = start.id
and
edge.end = end.id;
```

- Assigning the result of the SQL query to variables in a programming language, tuple by tuple, and building a structure for the nodes (e.g., a linked list):

```
// create an empty list
// 'edgel';
OPEN road;
WHILE (SQLCODE == 0)
| FETCH road INTO :
edge.start.x,
:edge.start.y, :edge.end.x,
:edge.end.y ; // add the
// variable edge to edgel |
```

CLOSE road;

- Processing the linked list in a function "buffer zone" that was previously implemented in a standard programming language such as C:

```
| roadBuffer = buffer (edgeL, 100); |
```

Experience shows that these incompatibilities cannot be completely hidden at internal and user interfaces, making the programmers’ tasks and, ultimately, the system users’ tasks, more difficult. This is an important consideration, because teaching and training costs for new systems are considerable parts of the overall costs of introducing a new information system.

Object-Orientation

The basic idea of object-orientation is the observation that the world is often perceived as consisting of objects, which interact in specific ways. The interaction among the objects can be seen as a command, or message, given to an object—either verbally or by some action such as physical force. Based upon common operations that can be applied to the objects—or commands to which they respond—they are grouped into classes. This concept was originally introduced into programming as part of the simulation language SIMULA (Dahl and Nygaard 1966), and later found to be generally valid and applicable.

In software engineering, object-orientation has become a design method that focuses as a first line of structuring on modeling objects as humans perceive them in reality. Unlike
previously used approaches, it combines modeling of the structure and the behavior of the objects. (Procedural abstractions model primarily the operations, while methods used for designing database schemas concentrate on the structure of the entities.)

The object-oriented method corresponds closely to the mathematical concept of multi-sorted or heterogeneous algebras (Birkhoff and Lipson 1970). From this point of view, the description of an object consists of a name for its type, a set of operations which are applicable to objects of this type, and a set of axioms which define the behavior of the operations. The important idea is that it is possible to write the axioms in terms of the operations, i.e., one defines the behavior of an operation in terms of other operations.

Object-Oriented Abstraction Mechanisms

This section introduces the notation of objects and the abstraction tools available to deal with them, following Dittrich's (1986) synthesis. A definition of object-orientation is that:

any entity, independent of whatever complexity and
structure, may be represented by exactly one object.

No artificial decomposition into simpler parts should be necessary due to technical restrictions. This is referred to as structural object-orientation.

Complex data types per se, modeling large objects such as entire cities (with all details about streets, buildings, etc.), do not overcome the problem of data structuring, and only the combination of complex object types and operations upon such instances provides the necessary view of objects. This second component of object-orientation is called operational object-orientation and requires that:

operations on complex objects are possible without having to decompose the objects into a number of simple objects.

The third notion of behavioral object-orientation states that:

a system must allow its objects to be accessed and modified only through a set of operations specific to an object type.

The object-oriented data model is built on the four basic concepts of abstraction (Brodie 1984): classification, generalization, association, and aggregation. Entity-Relationship-like diagrams for geographic examples will be used to visualize the abstraction mechanisms.

Classification

Classification is the mapping of several objects (instances) onto a common class. The word object is used for a single occurrence (instantiation) of data describing something that has some individuality and some observable behavior. The terms object type, sort, type, abstract data type, or module refer to types of objects, depending on the context. In the object-oriented approach, for every object there exists at least one corresponding class, i.e., every object is an instance of a class; therefore, classification is often referred to as the instance-of relationship.

A type characterizes the behavior of its instances by describing the common operators that can manipulate those objects (O'Brien, Bullis and Schaffert 1986). These operations are the only means to manipulate objects. All objects that belong to the same class are described by the same properties and have the same operations. For example, the model for a Town may include the classes Residence, CommercialBuilding, Street, and LandParcel. A single instance, such as the building with the address "26 Grove Street," is an object of the corresponding object type, that is, the particular object is an instance of the class Residence. Operations and properties are assigned to object types; for instance, the class Residence may

Egenhofer/Frank/URISA Journal 7
have the property NumberOfBedrooms, which is specific for all residences. Likewise, the class Street may have an operation to determine all adjacent Parcels (Figure 2).

Differences between objects of the same class are based upon their property values. Property values describe the individual characteristic of each object. For example, two LandParcels may be distinguished by their addresses, different values of their areas, or specific LandUseTypes.

**Generalization**

Generalization (not to be confused with the term used in cartography) groups several classes of objects with common operations into a more general superclass (Dahl and Nygaard 1966; Smith and Smith 1977; Goldberg and Robson 1983). The term superclass characterizes this grouping and refers to object types which are related by an is-a relation. The converse relation of superclass, the subclass, describes a specialization of the superclass. Frequently, the terms parent and child are also used for superclass and subclass, respectively. Though this terminology is helpful to clarify the dependency of subclasses from superclasses, it is not accurate with respect to the abstraction, because the relationship between parent and child is not is-a. Subclass and superclass are abstractions for the same object and do not describe two different objects. For example, each Residence is a Building; Residence is a subclass of Building, while Building is its superclass (Figure 3). The residence with the address "26 Grove Street," for example, is simultaneously an instance of the classes Residence and its superclass Building.

Two properties of generalization should be mentioned in more detail:

- A superclass may encompass multiple subclasses. For example, besides Residences, there may be other building types such as Hospitals and CommercialBuildings (Figure 4).
- Generalization may have an arbitrary number of levels in which a subclass has the role of a superclass of another, more specific class. For example, the specialization from Buildings to Residences can be extended with the classes RuralResidence and UrbanResidence, both being subclasses of Residence. While Residence is a subclass of Building, it is at the same time a superclass for RuralResidence and UrbanResidence (Figure 5).

**Association**

An association relates two or more independent objects and the relationship between objects is considered a higher level set.
object (Brodie 1984). The term set is used to describe the association, and the associated objects are called members. Hence, this abstraction is referred to as the member-of relation, but is also often called grouping or partitioning. An example of an association in the GIS domain is neighborhood, which relates a land parcel with its adjacent house lots.

The details of a member object are suppressed and properties of the set object are emphasized. An instance of a set object can be decomposed into a set of instances of the member object. Association applied to objects (members) produces a set data structure. Operations over sets are normally operations that are repeated for each member of the set. The implementation of the repetition may be a FOR EACH loop structure, as found in some modern programming languages, e.g., CLU (Liskov et al. 1981).

Aggregation

An abstraction mechanism, similar to association, is aggregation which models composite objects, i.e., objects which consist of other objects (Smith and Smith 1977). Several objects can be combined to form a semantically higher-level object, called aggregate or composite object, where each part has its own functionality. The terms subpart or component refer to the parts of the composite object. Operations of aggregates are not compatible with operations on parts, and vice-versa. When considering the aggregate, details of the constituent objects are suppressed. Every instance of an aggregate object can be decomposed into the instances of the corresponding component objects.

The relation established by aggregation is often called the part-of relation since aggregated instances are parts of the aggregate; the relationship converse to part-of is called consists-of. For example, a City may be modeled as an aggregate of all HouseLots, Streets, and Parks—they are part-of a City or, conversely, a City consists-of them (Figure 6).

Aggregation applied to objects (components) produces an aggregate (or record) type data structure. An operation over an aggregate consists of a fixed number of different operations in sequence or in parallel, one for each component. Hence, aggregation relates to sequence or parallel control structures.

Formalizations of Behavior

The literature concerned with object-orientation has evolved considerably over the last few years and the initial proposals for formalizations of complexly structured objects (Batory and Buchmann 1984) have progressed to detailed descriptions of the tools to model the behavior of objects. Since there is still a lack of consensus among different researchers in the field of object-oriented modeling about some details of these concepts, this section will attempt to synthesize the mainstream ideas of enhancing the semantic model with the crucial concepts of inheritance and propagation. The first form describes the derivation of properties in generalization hierarchies, while the second one deals with values in aggregation hierarchies. Sometimes propagation is also called upward inheritance (Barrera and Buchmann 1981; Brodie and Rlijanovic 1984), but it should become clear from this paper that these are two different concepts which must be separated.

Inheritance

In a generalization hierarchy, the properties and methods of the subclasses depend upon the structure and properties of the superclass or superclasses. Inheritance is a method to define a
that a subclass may inherit only parts of the operations prescribed by the superclass. Otherwise, complex exception rules apply (Borgida 1988).

Inheritance is transitive, i.e., the properties are passed along from a superclass to all related subclasses, and to their subclasses, etc. This concept is very powerful, because it reduces information redundancy and maintains integrity (Woelk and Kim 1987). Modularity and consistency are supported since essential properties of an object are defined only once and inherited in all relationships in which it takes part.

**Single Inheritance.** Inheritance can be strictly hierarchical; it is then often referred to as single inheritance. Single inheritance requires that any class has at most one single immediate superclass. This restriction implies that each subclass belongs to a single hierarchy group and that one class cannot be part of several distinct hierarchies.

The following example shows inheritance along a generalization hierarchy (Figure 7). RESIDENCE is the general superclass and URBANRESIDENCE and RURALRESIDENCE are the specific subclasses. All properties and operations of the class RESIDENCE are inherited by its two subclasses. All properties and operations of the class RESIDENCE are inherited by its two subclasses. For example, Resident and MovingIn are associated with the class RESIDENCE and inherited by all URBANRESIDENCES and RURALRESIDENCES. They are compatible with URBANRESIDENCES and RURALRESIDENCES. On the other hand, the operations defined specifically for a subclass are not applicable to objects of the superclasses. For instance, NextSubwayStop is a property which applies only to URBANRESIDENCES.

The transitive property of inheritance implies that any property is passed not only from the superclass to the immediate subclasses, but also to their sub-subclasses, etc. For example, the properties of a BUILDING, such as address and owner, are inherited by the subclass RESIDENCE, and also transitively to the sub-subclasses RURALRESIDENCE and URBANRESIDENCE (Figure 8).

**Multiple Inheritance.** The structure of a strict hierarchy is an idealized model and often fails when applied to real world data. Most "hierarchies" have a few non-hierarchical exceptions in which one subclass has more than a single, direct superclass.

---

**FIGURE 7.** Inheritance of the property resident along the generalization hierarchy.
Thus, pure hierarchies are not always the adequate structure for inheritance. Instead, the concept of *multiple inheritance* (Carl deli 1984) permits one to pass properties from several higher-level classes to another class. This structure is not hierarchical, because—in terms of the parent-child relation—one child can have several parents. In the simplest case of multiple inheritance, a subclass inherits properties from two distinct superclasses. For example, the different roles of a *LandParcel* as a *TaxableItem* and a *RealEstateObject* can be effectively modeled by multiple inheritance (Figure 9). A more complex GIS example shows how multiple inheritance combines two distinct hierarchies (Figure 10). The first hierarchy is determined by the separation of *TransportationLinks* into *ArtificialTransportationLinks* and *NaturalTransportationLinks*. Highways and Channels are *ArtificialTransportationWays*, and *NavigableRivers* are *NaturalTransportationLinks*. Waterbodies with *Ponds*, *Channels*, and *Rivers* form a second hierarchy, in which two types of rivers are distinguished: *NavigableRivers* and *UnNavigableRivers*. Classes with properties from both hierarchies are *Channels* that are *ArtificialTransportationLinks* and *Waterbodies*, and *NavigableRivers* that are *Rivers* and *NaturalTransportationLinks*. These hierarchies cannot be compared with each other, because a *Waterbody* is not necessarily a *TransportationLink and vice-versa, not every TransportationLink is a Waterbody either; however, the hierarchies share common subclasses because *Channels* are both *Waterbodies* and *ArtificialTransportationLinks*, and *NavigableRivers* are *Rivers* and *NaturalTransportationLinks*. Other classes, such as *Highway* or *Pond*, belong only to a single hierarchy in this schema.

The problem of *name clashes or inheritance conflicts* has received much attention. If a class

**Figure 8.** Transitive inheriting the properties address and owner to all subclasses of *Building.*

**Figure 9.** A *LandParcel* modelled with multiple inheritance as a *TaxableItem* and a *RealEstateObject.*

**Figure 10.** A GIS example of the use of multiple inheritance.
has several superclasses, it may inherit distinct operations with the same name, but different meanings. For instance, a \texttt{LANDPARCEL} has a value as a \texttt{REAL ESTATE OBJECT} and as a \texttt{TAXABLE ITEM}. Both values are based on different assessments and used for different purposes. Single inheritance has a simple rule to resolve such name conflicts: it gives preference to the most specific method (i.e., the one associated with the most detailed superclass). This selection may not necessarily be what was intended with the model, but it is at least a simple and consistent rule. For multiple inheritance, there are no such simple conceptual rules that could capture the intended meaning. Frequently, the conflict is resolved by giving preference to the methods in the order they are listed in the data definition (Stonebraker and Rowe 1986); however, this would not be a valid solution for the value of the \texttt{LANDPARCEL}. Since the two names actually describe two different properties, it is necessary to distinguish between them, e.g., by tagging the property names with their class names, e.g., \texttt{REAL ESTATE OBJECT.value} and \texttt{TAXABLE ITEM.value}.

Inheritance for GIS Modeling. Usually, a GIS contains many application-specific classes of objects such as \texttt{CITIES}, \texttt{RIVERS}, \texttt{ROADS}, \texttt{BUILDINGS}, \texttt{HOUSEOWNERS}, \texttt{PARCELS}, \texttt{SOILS}, and their detailed subclasses. A number of operations is associated with each class. For example, owners may sell their land, a new road may be constructed, or a building may be demolished. Some of these operations are similar, e.g., the operation to determine all \texttt{CITIES} that lie within a \texttt{COUNTRY} and the operation determining all \texttt{BUILDINGS} within a \texttt{CITY} can be both interpreted as the geometric operation inside (Egenhofer and Herring 1990). Identifying and describing formally similar objects with common operations forms one of the goals of conceptual modeling, which allows the users’ models of their mini-worlds to be implemented with the least redundancy.

One part of an application model is the definition of a set of classes as the abstraction of objects with common properties. For each class the appropriate operations and relationships must be defined, including operations that combine objects of different classes. For example, the class \texttt{BUILDING} has the operation \texttt{onParcel}, which checks whether a building is located inside a parcel. Since inside applies also to many other objects, such as \texttt{CITIES} with respect to \texttt{COUNTRIES}, many similar, often highly redundant operations are defined and implemented which make modifications difficult and yield frequent inconsistencies.

Inheritance is an extremely effective means to model such situations in a GIS, formalizing the structure and properties of the object classes. By the definition of a general superclass for each specific concept, common properties may be defined in a single high-level class and inherited to the classes of the GIS application. For example, the superclass \texttt{GEOMETRIC} defines geometry with properties, e.g., location, spatial relationships such as neighborhood, inclusion, intersection, distance, and direction (Egenhofer and Herring 1990). A class in the user model can be defined as a subclass of \texttt{GEOMETRIC} inheriting all these properties. For example, the class \texttt{BUILDING} is a \texttt{SPATIAL} object. \texttt{BUILDING} can be described as the subclass of \texttt{GEOMETRIC} inheriting all spatial properties (Figure 11). Other properties can be defined in a similar way. For example, database properties, such as persistency, multi-user access, and transaction control, can be inherited from a superclass \texttt{PERSISTENT}. The general database operations, such as store, delete, retrieve, and modify, are then defined for the class \texttt{PERSISTENT} and passed to the specific object classes. If the class \texttt{BUILDING} is a \texttt{PERSISTENT} class, \texttt{BUILDINGS} can be stored, deleted, retrieved, and modified.

It is obvious that this type of modeling requires multiple inheritance (Frank 1988b). A class can have a multitude of diverse

![Figure 11](image)

The class \texttt{BUILDING} inherits the geometric properties from the superclass \texttt{GEOMETRIC}.
properties to be inherited. Important GIS properties are Persistent providing database behavior, Geometric inheriting common geometric concepts, Graphical providing graphical display, and Temporal for the description of the history of data (Frank and Egenhofer 1992; Frank 1988b). For example, buildings with geometric, graphical, and database behavior can be modeled by creating a class Building with properties, such as address and owner, and then inheriting Geometric, Graphical, and Persistent properties from the corresponding superclasses (Figure 12).

Propagation

Frequently, complex objects are not independent and have some property values which rely upon values of other objects (Kim et al. 1987). In aggregation hierarchies, for example, some values of a composite object depend on values of the properties of its components. These dependencies are meaningful and in order to guarantee consistency and integrity, their correct modeling is crucial. Of course, a composite object may have property values which it owns specifically and which are independent from those of their components. In contrast to less powerful models which require redundant storage of such values, the object-oriented model allows objects to have properties with values which rely on values of other objects and models these dependencies consistently. This model is superior, because it enforces integrity by constraints. These derived values frequently describe geometric or statistical properties. Particularly in GISs, a large number of attribute values at one level of abstraction depends upon values from another level and must be derived from them. When combining local and regional data, this concept of modeling data at different levels of resolution must be used to furnish consistency among dependent values. The population of a county, for example, depends on the populations of all related settlements; therefore, the value for the property population of a county must be derived from all values of the property population owned by the settlements (Figure 13).

While inheritance describes properties of subclasses (types and operations), propagation describes how a value of a property of one class is derived from values of properties of another class (Egenhofer and Frank 1986). The notion of propagation is sometimes also used for modeling the behavior of operations, such as copy, destroy, print, and save, upon composite objects and how these operations propagate to their components (Rumbaugh 1988), and consistency of actions (Ellis et al. 1993). Here, propagation describes dependencies in the reverse direction—from the components to the composite object. Formal definitions of propagation, also demonstrating the differences between inheritance and propagation, have been given in terms of first-order predicate calculus (Egenhofer and Frank 1989; Egenhofer and Frank 1990) and are also part of more comprehensive algebras for complex objects (Shaw and Zdonik 1989; Beeri and Kor- natzky 1990). There may be multiple methods to deduce a value; a value may also be re-

FIGURE 12.
The class Building inherits the various properties from multiple superclasses.
The value of the County population is propagated as the sum of the population of the aggregated Settlements.

![Diagram showing Settlement and County with population values and a sum arrow]

recorded explicitly even if it could be derived from others. It may be necessary to resolve discrepancies and errors resulting from such "redundancies," e.g., by assessing the quality of the different results.

Propagation becomes trivial if the complex object happens to be composed of a single part and the value of the aggregate refers to a single value of the part; however, in most cases propagation involves values of multiple components. If more than a single value contributes to the derived value, the combination of the values must be described by an aggregate function. Aggregate functions combine the values of one or several properties of the components to a single value. This value reduces the amount of detail available for a complex object. It may determine the sum or union of values of the components, or define a specific, outstanding part such as the greatest, heaviest, or conversely, the smallest or lightest one. On the other hand, it may be representative such as the average or weighted average of the values of a specific property. Common aggregate operations are minimum, maximum, sum, average, and weighted average. For example, the population of the biggest city in a county is the maximum of the populations of all its cities; the area of a state is the sum of the areas of all its counties; the population density of the state is the average of the population density of its counties weighted by the county areas.

The concept of propagation guarantees consistency, because data are only stored once and the dependent values of the aggregate are derived; therefore, derived aggregate values need not be updated every time the components are changed. Of course, updates underlie the common rules for updates of views (Dayal and Bernstein 1978), i.e., no derived properties can be updated explicitly, but only the fundamental properties. For example, modifying the population of Penobscot County by assigning the value 65,000 to the county population if the town population of Orono grows by 5,000 is not allowed. Instead, the population of the settlements must be modified which implicitly updates the county population.

Two characteristics of propagation are observed: (1) the propagation of an aggregate value may involve several values from different classes, and (2) propagation may be transitive, i.e., propagated values may be used to derive further aggregated values. The following two examples clarify these characteristics. The area of a County depends on the areas of its LandParcels, Roads, Lakes, and Rivers and must be derived as the sum of all areas of these components. An example for the transivity of propagation is the population of the largest county (PopulationOfLargestCounty) in a State, which depends on the population of the Counties, which in turn is the sum of the populations of their Settlements. Implementation considerations recommend that these computationally expensive aggregate operations are reduced to a minimum to improve query performance (Blakeley, Larson and Tompa 1986), sometimes by introducing redundant storage of aggregated values.

Alternatives to Object-Oriented Data Models

With the knowledge and understanding of the object-oriented model it is now possible to assess its benefits and trade-offs when looking at other data models.

The Relational Data Model

Today, the relational data model is the most common one. It seems to be suitable for modeling commercial data for which humans may have the mental model of tables such as bank accounts; however, it is too simplistic for modeling data that
describe spatial phenomena. Though geographic data may be modeled with strings of characters, reals, and integers in table format (Osborn and Heaven 1986; Waugh and Healey 1987; Abel 1988; von Meyer 1989; Herring 1992), a higher-level model is desired in which more complex data types can be used (Nyerges 1980; Egenhofer and Frank 1987). The relational model lacks, for instance, the powerful concept of recursion, which is crucial for modeling multi-scale situations such as spatial data and their subdivisions. Regions can be decomposed into several smaller regions each of which may be continuously decomposed further. This recursive decomposition applies, for instance, to the subdivision of land parcels. Likewise, existing commercial database management systems are not sufficient for applications of non-standard database applications (Härder and Reuter 1983) such as geographic information systems or automated cartography. Each of these applications contains substantial amounts of "real world" data with geometric aspects. Their composition is too complex to be efficiently managed in conventional database management systems. Current database management technology falls short of the requirements for engineering (Sidle 1980; Udagawa and Mizoguchi 1984; Buchmann and de Celis 1985) and scientific (French, Jones and Pfaltz 1990) database applications, because the treatment of complex objects (Lorie and Plouffe 1983), such as spatial objects in GIS/LIS (Frank 1984), circuits in VLSI, or molecules in chemistry (Bastry and Buchmann 1984), is not supported and appropriate mechanisms for data structuring are missing. As a consequence, performance is unacceptable when a database is populated with large amounts of data (Wilkins and Wiederhold 1984; Härder and Reuter 1985; Maier 1986). When applied to geographic data handling, the same problems have been observed (Smith and Frank 1990; Guenther and Buchmann 1990).

Extensions of Traditional Models

A number of methods to capture more semantics in the data model have been proposed in the literature, often referred to as semantic data models (Hull and King 1987). Most of these methods have not been implemented, and only a few of them, such as SIM by Unisys, are readily available. It is suspected that if the transition from the first scientific publications to commercial products happens at the same pace as with relational technology, then the first commercial products based upon semantic data models may become reality in the mid '90s (Peckham and Maryanski 1988).

In the past, considerable effort was made to enrich existing data models with facilities to treat complex objects. ADT-INGRES (Stonebraker, Rubenstein and Gutman 1983) extends the relational model with features to define more complex types. DAPLEX (Shipman 1981) is a functional language that includes hierarchical relationships and transitive closure. The extension of the relational model with a surrogate concept provides hierarchical relationships between relations (Meier and Lorie 1983). The NF2 model (Schek 1985) supports composite attributes being tuples or relations performing a hybrid of the relational and hierarchical data models. The POSTGRES data model (Stonebraker and Rowe 1986) is based on the relational data model and ties complex objects and the corresponding operations close together. The molecule-atom data model (MAD) and the molecule algebra (Mitschang 1989) extend relational algebra to model aggregates as complex objects. Investigations of how geometry could be modeled using these extended data models showed that only partial remedies are provided with these extensions (Kemper and Wallrath 1987), because they lack sophisticated object-orientation.

Only recently has the design of database management systems specifically for spatial information systems become a research topic (Frank 1981; Lipeck and Neumann 1986; Schek and Waterfall 1986; Smith and Frank 1990) and a few experimental spatial database management systems exist, such as PANDA (Frank 1982), or database management systems with extensions for spatial data handling, such as DASDBS (Waterfall, Wolf and Horn 1988). Current commercial GIS software systems tend toward sophisticated software engineering methods (Aronson and Morehouse 1983) and object-oriented concepts for geometric data handling (Herring 1987;
Chance, Newell and Theriault 1990); however, database management systems, and especially object-oriented ones, have not yet been widely incorporated into commercial GIS systems.

Conclusions
This paper presented a synthesis of an object-oriented data model based on the abstraction concepts of classification, generalization, association, and aggregation. Inheritance and propagation are crucial for modeling the behavior of complex objects. It has been demonstrated how object-oriented modeling can serve as a suitable method for the design of spatial information systems. Current data models and database technology are not sufficient for the specific tasks of dealing with large amounts of spatial data. Recently, research in non-standard database environments promoted an object-oriented model, which promises to overcome some problems that make conventional database management systems unsuitable, such as the lack of modeling power to adequately describe complex objects and the unacceptably slow performance of current implementations. Recent benchmarks (Cat tel and Skeen 1992) of some object-oriented database management systems have given evidence that they can perform by more than an order of magnitude better than their relational predecessors.

Object-orientation is an encompassing methodology that can be applied to all stages of designing and implementing a large, complex software system. Commercial products for object-oriented database management systems have become available (Maier et al. 1986; Fishman 1986), with more to appear on the market in the near future (Deux et al. 1991; Lamb et al. 1991). Future GISs are excellent candidates for applying object-oriented concepts in order to improve their design, implementation, and maintenance. All three object-oriented concepts—modeling, software engineering, and database management systems—are important for GIS personnel:

- The object-oriented abstraction mechanisms are necessary to model the complex situations, such as geometric objects, which can change over a period of time. The complexity of spatial objects requires methods to define and use appropriate spatial data types and operations. The object-oriented model supports this task with appropriate abstraction mechanisms. In particular, the powerful concept of inheritance gives rise to consistent and concise definitions for properties such as geometry, graphics, and persistency. Data structures are necessary for recursive object definitions, such as areas being subdivided into other areas, and transitive closure operations.

- Object-oriented programming languages will be needed to implement the future GIS most efficiently. In recent years, research in software engineering has promoted an object-oriented design method by which real world objects and their relevant operations can be modeled in a program. This approach is most useful for application areas like GISs because it naturally supports the treatment of complex, in this case geometric, objects (Kjerne and Dusker 1990). Compared with conventional data models, an object-oriented design is more flexible and better-suited to describe complex data structures.

- Object-oriented database management systems must be used to exploit the modeling power and performance to manage and retrieve spatial data. Spatial information systems will benefit from the use of object-oriented database management systems in various ways: (1) The architecture of a GIS will become clearer such that the maintenance of GIS software will be easier and its life cycle will be longer. (2) Programmers should not worry about aspects of the physical location of data; instead, a unified set of commands provides the functionality for storage and retrieval of data. (3) By using a database management system, data are treated by their properties; the object-oriented approach groups these properties into possibly complex objects and corresponding operations.

Acknowledgement
Our understanding of object-orientation has been improved during many discussions with Renato Barbosa and Alex Buchmann. John Isner made us familiar with the object-oriented concepts in C++. Robert Cicogna, Doug Hudson, and Jeff Jackson read and commented on an earlier version of this paper.

References
Al-Tala, K. and R. Barrera. 1990, "Temporal Data and GIS: An Overview." In Proceed-
ings of GIS/LIS '90, Anaheim, CA. 244–254.


Benefit-Cost Principles for Land Information Systems
Gregory L. Poe, Richard C. Bishop and Jeffrey A. Cochrane

Gregory L. Poe is a doctoral candidate in natural resource economics in the Department of Agricultural Economics, University of Wisconsin-Madison. His research interests include management of environmental risk and externalities, non-market valuation, and evaluation of natural resource and environmental policies.

Richard C. Bishop is a professor in the Department of Agricultural Economics, the Center for Resource Policy Studies, and the Institute for Environmental Studies at the University of Wisconsin-Madison. His duties include teaching in environmental and resource economics and benefit-cost analysis.

Jeffrey A. Cochrane is a doctoral candidate in resource economics in the Department of Agricultural Economics, University of Wisconsin-Madison. His research interests include exploration of interdisciplinary social theories to understand collective decision-making with respect to resources.

Abstract: Past investigations into the economics of land information have focused on the cost aspects of alternative systems, with the benefits perceived as intangible. This paper reviews the economic concepts that relate to the valuation of information systems and demonstrates that cost comparisons generally provide a biased estimate of the net social benefits of improved land information systems. For this reason, application of the avoided cost approach could lead to misguided advice about which systems are most desirable from an economic perspective. Possible methods for more accurate estimation of the net social benefits are suggested.

With some notable exceptions (e.g., Blaine and Randall 1987; WUNDERLICH and Moyer 1984) past investigations into the economics of land information systems have apparently concluded that information economics has little to offer in the evaluation of alternative information systems. These investigations then proceed either to conduct cost comparisons of alternative information systems using avoided cost techniques, or to discuss in broad terms the 'intangible' aspects of information and information systems (e.g., Epstein and Duchesneau 1990; GURDA et al. 1987; Moyer and Niemann 1990).

Note: This project was supported by the College of Agricultural and Life Sciences, University of Wisconsin-Madison, and funding was provided, in part, by the U.S. Department of Agriculture Soil Conservation Service through the CONSOIL project.

The objective of this paper is to review the economic concepts that relate to the valuation of information systems and to suggest principles for the proper accounting of benefits and costs of improved land information systems. The motivating assumption is that the economic potential of land information systems cannot be adequately addressed until the underlying economic foundations are understood.
Particular emphasis is given to the demand characteristics and value of information, an area of study that has received much attention in the general economic literature but is notably limited in studies concerned with the benefits and costs of land information.

The prevailing view in the land information economics literature is that the avoided cost approach underestimates the value of improvements in land information system services (Blaine and Randall 1987; Epstein and Duchesneau 1990). A careful examination of the economics of land information provision will show that either underestimation or overestimation of the "true" benefits will occur when the avoided cost approach is applied. Possible methods for more accurate estimation are suggested.

Concepts in Land Information Economics

The objective of applied information economics is to identify the information system that will maximize net social benefits over time (Eisgruber 1973; Marschak 1968). This section develops a simple static graphical approach to evaluating the economic desirability of alternative information systems, and demonstrates that the avoided cost approach will, in general, provide a biased estimate of the net benefits of alternative information systems.

A 'Market' for Information

Information can be viewed as an intermediate product in an inquiry- and decision-making process (Chavas and Pope 1984; Marschak 1968). It is the output of a data system or inquiry process that organizes, stores and transmits data, and it serves as an input into decision-making processes. In this manner it is possible to envision a land information market that incorporates both 'supply' and 'demand' characteristics.

Our conceptualization of this land information market focuses on land information that is publicly provided and has public good characteristics, factors that make land information quite a bit different from conventional commodities (e.g., oranges, cars, etc.) wherein market supply and demand curves represent consumer and producer quantity responses at different prices. This distinction is important for two reasons. First, the quality of publicly provided information is determined exogenously by policy makers, information specialists, land conservation committees and so on. Thus, it is likely that the level of information provision will deviate from 'market clearing' or 'equilibrium' conditions where the marginal social costs of producing information equal the marginal social benefits. Similarly, the price charged for information is not necessarily related to the marginal cost of provision. It is typically the case that public land information fees are set administratively, not by market interactions.

Second, the public goods characteristics of information necessitate a different method of conceptualizing a market than conventional analyses of supply and demand (Samuelson 1954; Boadway and Wildasen 1984). For most goods and services, the aggregate demand represents the horizontal summation of all individual demands. For public goods, however, the level of the good is fixed and individuals instead attach a marginal benefit or a willingness to pay for the good, a value which may vary considerably across individuals. The aggregate 'demand' for public goods represents the vertical summation of the marginal benefits attributed to the service by all individuals.

Within this framework, a hypothetical market for information can be graphically depicted (Figure 1). In this figure, information is represented on the horizontal axis by the inverse uncertainty parameter, 1/θ. The parameter θ might be interpreted as the standard deviation of the distribution of a random variable. A distribution with a smaller standard deviation is said to provide more information about the random variable. The vertical axis in Figure 1, which is expressed in monetary units, measures the marginal benefits and costs of additional improvements of information.

In Figure 1 the marginal benefits or maximum willingness-to-pay curve is defined as the vertical summation of individual bid/valuation curves for incremental improvements in information. It is assumed that the willingness to pay for additional units of information is characterized by diminishing returns: initial reductions in uncertainty are valued highly by individuals, while subsequent units have
Marginal benefits curve for additional tests. Using similar logic, and noting that each subsequent test costs the same but has a lower impact on reducing uncertainty, provides a rationale for the rising and convex marginal cost curve.

**Benefit-Cost Analysis for Land Information Systems**

Across disciplines benefit-cost analysis (BCA) has taken on a variety of interpretations. For economists, BCA is the application of the tools of welfare analysis to evaluate how a particular investment or monetary decision affects economic efficiency or net social benefits. The area underneath the marginal benefits curve in Figure 1 provides a monetary measure of the total benefits associated with each subsequent improvement in information. Similarly, the total costs for each level of information are measured by the area beneath the marginal cost curve. The difference in the area under the marginal benefits curve and the area below the marginal cost curve defines economic surplus—a monetary measure of social welfare. For example, A + B denotes the economic surplus at the equilibrium value of $1/\theta^*$ in Figure 1. If, instead, only $1/\theta^A$ was supplied, then the surplus value would only be A. Similarly, if information were supplied at level $1/\theta^c$, then the surplus value would be A + B - C. Clearly, surplus is maximized where marginal benefits equal the marginal costs of information.

Benefit-cost analysis measures how the aggregate surplus value...
changes with new policies or projects. These changes in surplus are called the net benefits, or simply the benefits, of the project or policy change. Basically, the criterion for project acceptance is that there is a net gain in economic surplus resulting from the project (Broadway and Bruce 1984; Just, Huetth and Schmitz 1982; Anderson and Settle 1977). In practice, this is measured by subtracting the change in total costs from the change in total benefits.

The prevailing technique used in evaluating improvements in land information systems is based on the following avoided-cost philosophy:

*The benefits generated from a government operation can be represented by the costs avoided as the result of the operation. These savings are properly interpreted as benefits. The rationale, in terms of demand and expenditure, is that one would be willing to pay an amount equal to the cost savings in order to obtain the savings.*

(Epstein and Duchesneau 1984, p. 7)

While relatively simple and easily applied in an accounting framework, the avoided cost approach only provides an exact (or ‘true’) measure of changes in surplus under the very restrictive assumption that the level of information services is fixed by the public sector both before and after a technological change. For example, in Figure 2, assume that a technological change results in a downward shift in the marginal cost of providing information from MC to MC'. If the quality of information remains at 1/θ both before and after this shift, then the avoided cost benefit measure, B, would exactly equal the change in surplus. Under these conditions, the avoided cost approach does provide a valid and complete measure of the net benefits of the change according to benefit-cost theory. In essence, this has been the motivation for the parity approach adopted by the Dane County Multipurpose Land Information System (MPLIS) (see Guda et al 1987).

It is more common, however, that shifts in the supply of information are accompanied by a change in the quality of information allocated by the public sector. Computerized methods, for example, may offer greater precision as well as reducing the cost of land information provision. In these instances the avoided cost technique will provide a biased measure of change in surplus; the direction and the magnitude of the bias will depend upon the relation between the equilibrium point (where marginal benefits equal the marginal costs of information) and the actual level of information provision. For example, if the level of information were set at 1/θ before the supply shift and 1/θ' afterwards, the true benefits measure would equal B + C + D. In contrast, the avoided cost benefit measure would correspond to area B + D. Thus, in situations where information provision is set at levels at or below the equilibrium, the avoided cost measure of benefits provides an underestimate of
the true benefits of improvements in information provision. The belief that this type of disequilibrium represents the situation in the real world appears to be the motivating factor for investing in many land information systems.

Quite a different result occurs if the level of information provision is set at levels exceeding the market clearing conditions. Such a case might at first seem somewhat counterintuitive, but could occur in instances where technological advances are so rapid that they exceed the needs of information users. Our observations of overzealous information technicians who are infatuated with the technology indicate that indeed this situation is not so unlikely. These conditions are demonstrated graphically in Figure 2 in a move from $1/\theta^2$ to $1/\theta^3$, where, in this case, $1/\theta^2$ is assumed to represent the initial conditions. Under these assumptions the true benefits are given by $B + D + E + F - H$, while the benefits as measured by the avoided costs approach are $B + D + F + G$. As such, the avoided cost approach provides an overestimate of the benefits of the new information system.

The implication of the above analysis is that, in cases where information is publicly provided, the avoided cost approach does not, in general, provide an unbiased measure of the ‘true’ benefits of an improvement in land information services. Knowledge of, or strong assumptions about, the demand for information and the structure of the information market are required before broad statements about the direction and degree of the biases associated with the avoided costs approach can be made.

In a similar vein, a second limitation of the avoided cost approach is that it may be an invalid measure of benefits in evaluating new types of information. The question of validity would arise when either the old system could not provide the new type of information at any cost or the costs of providing information with the old system exceed society’s ability to pay for that type of information. In the former case, there are simply no costs with which to make a comparison. In the latter case, the costs of providing the new information with the old system do not serve as a valid base for comparison. In this instance, cost savings using the old system’s costs to provide the additional information do not represent a true benefit to society. Instead, the correct theoretical measure of the benefits from previously unavailable information is the resulting increase in economic surplus.

### Decision Analysis

**Theory.** Decision analysis typically employs an expected utility framework incorporating Bayesian updating (Anderson, Dillon and Hardaker 1977; Lal-Valle 1978; Winkler 1972). Consider a simple example from Bquet, Halter and Conklin (1976). A pear farmer believes there is a chance that frost will hit on a particular night. The farmer must then decide whether or not to use costly heaters to protect the crop. Under the expected utility framework, the farmer is assumed to choose an action which maximizes expected utility.

Suppose there is frost and the farmer did decide to use heaters. The farmer is assumed to
know what level of income will thereby be achieved, and can assess the utility derived from that income. Suppose then that frost does not occur, but the farmer did still decide to use heaters. (Obviously, the decision is made before the farmer knows whether there will be frost.) The farmer can again calculate net income and utility under this scenario. The expected utility framework posits that this farmer compares the two possible utilities resulting from a decision to heat, multiplies them by the respective probabilities of frost, and adds them together. This total is the expected utility from a decision to heat. The process is repeated for a decision not to heat. The two expected utilities are compared, and the farmer then makes that decision which results in highest expected benefits.

Of course the farmer may wish to purchase information in the form of a weather forecast, knowing that the weather forecaster has access to considerably more data. Assume that the farmer has previous experience with weather forecasts, and knows that they enable better prediction of frost. More precisely, the farmer has an idea about the likelihood of a prediction of frost given that frost actually will occur. The farmer then must decide whether or not this “information service” (Hirshleifer and Riley 1979) that generates information is worth purchasing.

Under Bayesian updating, the farmer will utilize the weather forecast to update the assessment of expected utilities by combining the likelihood of a true forecast with prior beliefs about the probability of frost, using the well-known laws of probability derived by Bayes. By comparing expected utility with and without information, the farmer can determine the ex ante (before acquisition) value of an information service. This is sometimes termed the farmer’s “willingness-to-pay” or “bid price” for an information service.

**Application.** While interesting from a theoretical standpoint, calculation of the bid price for information via decision analysis is extremely data intensive. It requires the elicitation of subjective prior probabilities for all the possible states of the world (in the orchard example there were only two), knowledge of the consequences of each action for each state (in the example there were four consequences to consider), direct elicitation of preferences and determination of utilities for each consequence (again, four), and elicitation of subjective likelihoods of true information messages (in the example there were two kinds of forecasts, times two states, yielding four likelihoods). This approach may be feasible only for very simple types of information to be used for very simple decisions. Moreover, there is some concern, expressed more frequently by non-economists, whether particular decision-makers behave in a manner consistent with Bayesian updating and the expected utility axioms.

Returning to the example of the farmer and the weather forecast, presumably the farmer’s assessment of the likelihood of a true forecast was based on previous experience with such forecasts. The decision to purchase a new forecast was based on the evaluation of the consequences of frost damage, not on any change in the likelihood of a true forecast. What if an entirely new type of forecast suddenly becomes available? For example, suppose that the expert now has access to satellite data, where previously she/he relied on ground-sensing devices. The farmer has no terms of reference with which to reassess the likelihood of a true forecast.

For more complicated problems, or for problems in which assessment of likelihoods is difficult, an alternative method entails the direct determination of the bid price of information through “contingent valuation” (Mitchell and Carson 1989). Essentially, this method employs personal or telephone interviews, or mail surveys to ask people about the values that they would place on non-marketed commodities “contingent on the existence of a market or other means of payment” (Anderson and Bishop 1986, p. 91). The farmer in the example could be asked to specify a bid price for a forecast that has a certain, presumably higher, likelihood of being true.

The contingent valuation method is increasingly accepted as a useful tool for benefit-cost analysis. It was authorized for the valuation of outdoor recreation in the *Economic and Environmental Principles and Guidelines for Water and Related Land*
Resources Implementation Studies (U.S. Water Resources Council 1983). Later, the U.S. Army Corp of Engineers prepared its own manual for applying the method (Moser and Dunning 1986) and has conducted many contingent valuation studies (Mitchell and Carson 1989, p. 13). The contingent valuation method was deemed acceptable by the U.S. Fish and Wildlife Service for human use and evaluation studies (U.S. Fish and Wildlife Service 1985). The U.S. Environmental Protection Agency, in its Guidelines for Performing Regulatory Impact Analysis, lists contingent valuation as one of the four basic methods for valuing the environmental benefits of proposed regulations (U.S. Environmental Protection Agency 1983, p. 9). With some caveats, the contingent valuation method was endorsed for estimating the damages done by releases of oil and toxic chemicals into the environment under procedures promulgated by the U.S. Department of the Interior (1986) to implement the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA) and amendments to the Clean Water Act. While contingent valuation has not, to our knowledge, been applied to the evaluation of land information systems or such other technologies, research to evaluate its potential might pay rich dividends.

In an application to land information systems, individuals might be asked their willingness to pay (bid price) for specific types of land information products. Individuals would be presented with an array of land information products from, say, a zoning map and separate topographic map printed to different scales requiring visits to separate offices in different parts of town, to a computer-generated map containing both zoning and topographic data overlain with satellite imagery of ground cover and sent via electronic data transfer directly to the consumer’s office. The potential success of contingent valuation in specific applications depends on the ability of survey respondents to imagine the purchase of a product outside of their experience. Significant progress has been made in recent years in assessing the reliability and validity of contingent valuation, and in the design of surveys to elicit the most accurate values (Mitchell and Carson 1989; Bishop and Heberlein 1990; Cummings, Brookshire and Schulze 1986).

Transactions Cost Analysis

Decision analysis requires direct determination of the demand for information. In contrast, transactions costs analysis involves measuring the effect of information on supply and demand of goods for which information is used. In this analysis, it is posited that market transactions are impeded by lack of information, a factor that increases the costs of voluntary exchange and reduces the aggregate welfare of society. In a seminal article, Coase (1960) noted:

In order to carry out a market transaction it is necessary to discover who it is that one wishes to deal with, to inform people that one wishes to deal and on what terms, to conduct negotiations leading up to a bargain, to draw up a contract, to undertake the inspection to see that the terms of contract are being observed, and so on. These operations are often extremely costly, sufficiently costly at any rate to prevent many transactions that would be carried out in a world in which the pricing system worked without cost. (p. 148)

Dahalan (1979) categorized these transactions costs, and relates all transactions costs to resource losses and imperfect information.

Both search and [market] transaction costs owe their existence to imperfect information about the existence and location of trading opportunities or about the quality or characteristics of items available for trade. The case is the same for bargaining and decision costs: these represent resources spent in finding out the desire of economic agents to participate in trading at certain prices and conditions. What is being revealed in a bargaining situation is information about willingness to trade on certain conditions, and decision costs are resources spent in determining the terms of trade are mutually agreeable. Policing and enforcement costs are incurred because there is a lack of knowledge as to whether one (or both) of the parties involved in the agreement will violate his part of the bargain. Therefore, it is really necessary to talk only about one type of transaction cost: resource losses incurred due to imperfect information. (p. 148)

In brief, three general forms of transactions costs are recognized: market information costs, contracting costs, and enforcement costs. A convenient way of remembering this taxonomy is the acronym ICE (Bromley 1986). It is important to note
that the type of information captured by this taxonomy is subtly different from that considered under decision analysis. In transactions cost analysis, information is viewed as a required input into the market process. With it, there will be a transaction; without it, there will not. Information here is akin to a transportation cost to ship a product to a retailer, or a telephone charge for service required to take orders from consumers. Note that probability does not enter into this type of analysis.

The ICE conceptualization is particularly relevant to the valuation of improved land information systems. Blaine and Randall (1987) argue that there is a demand, and thus a value, for improved market information.

The buyer is willing to pay for a certain amount of parcel quality information to help him identify the parcel which contains the characteristics he desires. In the traditional land market literature the implicit values of these characteristics are capitalized into the total value of the parcel. In a world of imperfect information, this capitalization will be imprecise and perhaps biased... Even if the implicit values of the parcel characteristics are perfectly capitalized into parcel values, the individual buyer is unable to observe either the implicit values or the characteristics. (p. 6)

Information of this type reduces costs to achieve established objectives, e.g., the purchase of a parcel with particular characteristics. With good quality information, the buyer may find the parcel quickly. Without good quality information, the

buyer may need to spend considerable time and money locating the parcel using low quality information. This subtracts from the buyer's ability to pay for the parcel once it is found, reducing effective demand for the parcel. This transaction cost is not distributed to anyone; it is lost. Society gains by reducing such costs.

Similarly, the contracting of soil and water conservation programs is facilitated by improved land information systems. For example, the computerized MPLIS in Dane County enabled contractors to prepare 20 percent of all conservation plans in Wisconsin, even though the county comprises only 4 percent of all the agricultural land in the state. This difference in rates is attributed to improved land information technology in Dane County (Moyer 1989; Licht 1989). There is also some evidence that improved land information services may aid in the enforcement of conservation mandates. Moyer and Niemann (1990) argued that the failure to share information layers has impeded enforcement of the "swampbuster provisions" of the 1985 Food Security Act in the prairie pothole region. Ventura (1988) demonstrated that a MPLIS can be used to target highly erosive parcels for conservation planning.

Using a land market for new housing development as an example, Figure 3 demonstrates a theoretical approach to valuing changes in transactions costs.

In this figure, D(.) represents the demand for land by housing de-

![FIGURE 3. Transactions Cost Method of Valuing Improved Information.](image-url)
velopers, a demand that is, in itself, derived from consumer demands for new housing units. The supply curve of land units represents the marginal cost of each additional unit of land. This supply curve, S(t,.) not only accounts for the opportunity cost associated with each parcel, but is also defined to include a unit transactions cost, t, associated with each unit of land purchased. These latter costs primarily consist of search costs for locating parcels that have suitable site characteristics, zoning patterns, current ownership patterns, soil and subsoil characteristics, location relative to major arteries, etc. Suppose that an improvement in information provision reduces these costs to \( t - r \) such that the new supply of land is specified by \( S(t-r) \). When all demand and supply effects are considered, the shaded area represents the benefit from an improved information system.

**Application.** As with the decision-analysis approach, transaction cost analysis poses formidable data-collection requirements. In this case the difficulty lies in determining supply and demand characteristics in all markets that might conceivably use the new information system. In each market, estimation of demand and supply would be required. In instances where various markets are related, estimation of cross-market effects might also be required.

If the “good” in question is an environmental good or natural resource such as erosion abatement, wetlands protection, or groundwater quality, the determination of demand and supply is particularly difficult. For these goods there are no demand or supply functions per se, but it is still possible to estimate marginal benefits and costs for specific levels of quality that could be used to guide public sector activity. Once again, contingent valuation may be the only feasible approach. In this case, the shaded area in Figure 3 would be estimated by asking “consumers” of, say, an A1 quality trout stream that has been negatively affected by agricultural runoff how much they would be willing to pay for better water quality, and presumably better fishing. These values would then be linked to improved information systems that would reduce runoff by facilitating more efficient farm planning. The state would, in a sense, be “supplying” better fishing quality via the information system. Aggregation across all “consumers” would yield the shaded area for the stream quality “market”, the total net benefit attributed to the improved fishing quality from improved information system. Aggregation across all affected markets besides the market for stream protection would also be required for a complete accounting of benefits.

**Case Study**

The principle aim of this paper has been to conceptualize benefits and costs of information, and to demonstrate, in a theoretical framework, that both benefit and cost characteristics should be considered in evaluating alternative information systems. This final section outlines how these theoretical principles might be extended to the proper accounting of benefits and costs in a case study.

**Project Description**

An interesting case in point which utilizes transactions cost analysis is that of the Land Conservation Committee (LCC) in Dane County, Wisconsin, which uses data from a number of sources, and processes the data using a MPLIS. Prior to the installation of a MPLIS at the LCC, these data were processed manually by visually comparing maps drawn to different specifications. The process of identifying and developing farm plans for those farms with erodible soils was highly labor intensive and time-consuming. The product of the MPLIS is a single-tract map overlain with soils data showing an erosion index and parcel ownership information.

The primary benefit of this system is perceived to be the acceleration of the contracting of conservation programs that require the development of farm plans. Secondary benefits may accrue to individuals if the information is made available to the public. These additional benefits might be analyzed using decision analysis as discussed previously.

At some point it may also be possible to link this MPLIS to satellite photographs identifying tillage practices and crop rotations. It would then be possible not only to identify erosion problems and facilitate the de-
The benefits that may be achieved under such a program stem from three primary sources:

1) Data processing cost savings,
2) Compliance enforcement cost savings, and
3) Increased compliance with the conservation restrictions resulting in lessened soil erosion.

Benefits from savings in the cost of data processing occur when the process of identifying erodible soils, notifying farmers, and developing farm plans is made more efficient. The costs of new hardware and staff training must be weighed against labor savings in processing data manually from soil and tract maps. Compliance enforcement is made more efficient if, by eliminating the need for visual inspection of farms, labor cost savings outweigh increased costs in remote sensing. The degree to which compliance will lessen soil erosion and the associated on-site costs (primarily present and future productivity losses) and off-site costs (instream damages—biological impacts, recreational impacts, water storage damage, navigation, and other "preservation values": off-stream damages—sediments in water conveyance, flood damage, water treatment) of erosion must be evaluated.12

For some policy applications it may be appropriate to weigh these benefits against any increased costs incurred by farmers in changing their practices, crop rotations, machinery requirements, and so on (e.g., Mueller, Klemme and Daniel 1985).

**Proper Accounting of the Benefits and Costs**

In the Dane County case, the extent to which the implementation of farm plans and compliance enforcement would have been undertaken by the LCC in the absence of an improvement in the information system must be considered before evaluating the change in costs. If it is determined that the LCC would have undertaken the process of planning and examining the same acreage even without this new system, then costs savings from the MPLIS will represent a valid partial measure of benefits to the county.

As Licht has demonstrated, however, counties that do not have computerized systems tend to complete many fewer farm plans. It is simply not feasible for counties to engage in such a massive program when labor costs associated with contracting are prohibitively high. If this is true for Dane County, the entire cost of the proposed MPLIS at the new, higher level of identification and planning should be subtracted only from those expenditures which would have truly been made under a smaller program in the absence of a MPLIS. Intuitively, given the high start-up costs of a MPLIS, it may be the case that cost savings in this area will be quite low or even negative if it can be shown that Dane County would not have been very enthusiastic about soil conservation in the absence of a MPLIS. Total benefits will then be total costs: the old, manual level of identification, planning and enforcement less the full costs of the proposed information system, plus benefits from increased production and lessened pollution, less any additional costs to the farmer. These considerations are summarized in Table 1.

Although this approach to benefit measurement will not completely capture the benefit gain associated with the movement along an erosion-abate-

---

**TABLE 1:**
Summary of the Economic Benefits of a Hypothetical MPLIS

<table>
<thead>
<tr>
<th>Description</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA PROCESSING COST SAVINGS = (Costs to process data to achieve the current or expected level of erosion identification and conservation planning under a manual system)−(Costs to achieve the new, high level identification and planning under the MPLIS)</td>
<td></td>
</tr>
<tr>
<td>COMPLIANCE ENFORCEMENT COST SAVING = (Costs under a manual system to process data, perform field inspections etc. to achieve the current level or expected level of enforcement)−(Costs under an MPLIS to achieve the new, high level of enforcement)</td>
<td></td>
</tr>
<tr>
<td>BENEFITS FROM COMPLIANCE = (Benefits from increased compliance resulting in greater production and less pollution)−(Increased costs to farmers resulting from changed farming practices as a result of compliance)</td>
<td></td>
</tr>
</tbody>
</table>
ment demand curve, it does provide a closer approximation to the true change in economic benefits than methods that only consider cost characteristics. This difference in estimates may have an impact on the decision to invest in information services.

Conclusion

Past studies have tended to focus on only the cost aspects of information systems, with benefits perceived as intangible. This paper has shown, in a theoretical framework, that such methods generally provide a biased estimate of the net social benefits of improved information systems and could lead to misguided advice about which information systems are most desirable from an economic perspective. Using concepts from welfare economics, a demonstration of this method was outlined in a case study.

Acknowledgements

We thank Jean-Paul Chavas, D. David Moyer and the anonymous reviewers for their helpful comments.

Notes

1. The approach taken here deviates considerably from the information-as-a-product approach taken in Blain and Randall and Wunderlich and Moyer. Here, land information is assumed to have public goods characteristics of non-rivalness in consumption and non-excludability. As an anonymous reviewer pointed out, most information systems contain elements of both public and private goods. For example, some public information systems (FDI, IRS) generate excludable, or highly restricted, information and some private information systems generate information with public attributes, such as multiple listings. As with monopoly and perfect competition in the theory of markets, the approach taken in this paper and that adopted in the information-as-a-product literature are two extremes in a continuum of possibilities. Interestingly enough, as will be discussed later in this paper, these extreme approaches provide a very similar critique of the avoided-cost method. Refer to Epstein and Duchesneau (1990) for further discussion about public good characteristics of land information systems.

2. In other words, this economic surplus is the amount of money that would just compensate society (including producers and consumers) not to consume information.

3. Blaine and Randall offer a market-based method for demonstrating that the avoided costs approach will underestimate the change in economic surplus when information is treated as a product and the information market is always assumed to adjust to a market clearing equilibrium. We note that some difficulties do arise with this approach in defining what constitutes a unit of information and for accounting for differences in quantity and quality of information. In spite of these difficulties, the Blaine and Randall presentation offers very strong insights about the relationship between the avoided cost and the 'true' benefits of improved information.

4. This can be presented in graphical terms using Figure 2, where instead the marginal cost curve, MC, would become vertical at some level of information. This perfectly inelastic section of the curve would indicate that the existing system could not reduce uncertainty beyond a certain point.

5. In economic theory utility is defined to be the level of well-being or happiness associated with a course of action.

6. Formally, the farmer calculates the bid price of information defined as

$$
\gamma = \max_a \left\{ \frac{E_n \left( u(y_s - \gamma; \pi, a) \right)}{a} \right\}
$$

where

- a = action (heat, don't heat)
- $y_s$ = income when state s occurs and decision a has been taken
- $\gamma$ = bid price of (willingness to pay for) information
- $\pi_m$ = posterior subjective probability of state s occurring given information m
- $u$ = utility.

The posterior subjective probabilities, $\pi_m$, are calculated using Bayes' Theorem from likelihoods and prior probabilities (Mood, Graybill and Boes), i.e., from the farmer's assessment of the likelihood of a true forecast and from the farmer's prior (without a forecast) beliefs about the probability of frost occurring. The expectation operator, $E_n$, is over possible states, while $E_a$ is over possible forecasts.

7. A separate survey design might also attempt to measure the option price of an information service which individuals presently do not use but have some non-zero probability of use in the future (see Luzar and Hanemann for a discussion).

8. Figure 3 represents a market for a traditional private economic good in which supply and demand reflect the aggregate quantity responses across individuals in the market. In contrast to previous graphs, these aggregate curves represent the horizontal summation of individual curves.

9. In this instance, the unit transaction cost is attached to the supply curve. Alternatively, with similar results, the transaction costs could be incorporated into an effective demand curve. In both cases the transaction costs drive a wedge between the seller's price and the buyer's willingness to pay. In actuality, it is unlikely that transaction costs will fall only on the producer or the consumer. Rather, a portion of the transaction costs will typically fall on both groups. For example, people selling land may factor the costs of a realtor to locate buyers, while purchasers may extend time and resources in search.

10. As Nicholson (p. 372) notes, this analysis of transactions costs is somewhat limited by its failure to consider agents that benefit from such costs such as middlemen brokers in real estate transactions. Discussion of such benefits is beyond the scope of this paper.

11. To this point, our analysis has focused on a land information market with a single use. The extension of this analysis to a MPLIS cannot be depicted in a single graph, but is, nevertheless, analogous to the single market case. For a MPLIS one would add up the benefits in all affected markets and compare them to the change in the cost of information provision.

12. Poe et al. develop a method for translating soil loss into economic values. See also Crosson and Stout, Clarc, Haverkamp and Chapman, and Ribaudo for discussions of the economic costs of erosion.
References


Eisgruber, L.M. 1973. "Managerial Informa-


Assessment of TIGER/Line Files for Rural Counties: The Case of Hays County, Texas

Ryan Rudnicki

Rudnicki received his Ph.D. degree from Penn State University. He is a professor in the Department of Geography and Planning at Southwest Texas State University. His most recent applied geography project consisted of building a county-wide digital database for Hays County, Texas.

Abstract: The TIGER/Line file system has been described as a major advance for digital mapping in the United States. These products have been greeted by mixed reviews, however. Examinations of the files for large metropolitan areas have disclosed several categories of errors and omissions.

Small cities and rural counties are at a disadvantage because they typically do not have the computer processing capability to generate maps based on TIGER. In addition, these places don’t have independently generated digital street maps of their jurisdictions to compare with the TIGER product. Thus it is difficult to ascertain whether the TIGER products are better or worse in rural areas than those generated for large metropolitan areas.

This paper examines the digital road and related attribute characteristics of the TIGER/Line files for a relatively small, rural county in Texas. The analysis gives other small counties and cities an idea of the kind and amount of scrubbing the TIGER files will need before they can be used with confidence by local government officials, departments, and agencies.

In 1987, the U.S. Census Bureau began releasing the TIGER/Line files. These files were a digital representation of a wide range of visible and non-visible linear features. Visible features included roads, pipelines and electric transmission lines, and streams and rivers for the entire United States. Non-visible items included lines representing various types and levels of census and political boundaries. The digital linear feature files were released on a state-by-state basis and are available by county. The first prototype file was released for Boone County, Missouri. Since 1987, the Census Bureau has released various versions of the files beginning with the precensus TIGER/Line files (1989), the Initial Voting District Codes TIGER/Line files (1990), and most recently, the post-census TIGER/Line files (1991). Consequently, the time is right for an assessment of TIGER/Line file quality because the database has reached its final form and can be considered stable for the next several years.

According to the Census Bureau, “these geographic/cartographic files are excellent as a base for a geographic information system (GIS)” (Bureau of the Census 1991, p. 5). Not only is good coverage of linear features provided, but also the Bureau claims that map feature locations approach or meet national map accuracy standards at a scale of 1:100,000. This paper evaluates the quality of the TIGER files as a base for a geographic information system by comparing maps produced from TIGER files with maps produced independently. Two TIGER/Line file components will be examined, road/street features and the attributes of those linear features. The former are cartographic elements; the latter, database elements.

What differentiates this study from others that investigated TIGER/Line file accuracy is the geographical focus. Other TIGER/Line analyses have been undertaken in large metropolitan areas: New York City and Houston for example (Hangan 1989; Barringer 1990). These places have computer resources at hand to help them extract various categories of information from the TIGER/Line files for analysis. These places represent a distinct minority of the total land area of the United States. Most of the nation does not consist of large metropolitan areas and the rural counties have not been heard from with regards to TIGER/Line file ac-
cacy issues. This paper attempts to redress that imbalance.

A difference between urban and rural TIGER/Line file accuracy and completeness stems from the origins of the TIGER/Line files for rural areas versus large urban areas. The TIGER/Line file database is derived mainly from two quite different sources. The first source consists of the DIME files from the 1980 census. These files are quite poor by most cartographic standards largely because the DIME file arcs (street segments) lack shape points. No matter how much a road may twist or curve, for the most part that street was represented by only two points: the begin-point and the end-point. In addition, the location of some DIME file begin- and end-points are misaligned giving straight roads on the ground a slightly zig-zag appearance on the map. Thus cartographic products based on the DIME files left much to be desired. These DIME files were moved into the 1990 TIGER/Line file framework without much change from their 1980 form. It was this product that the nation’s large cities have found to fall below expectations (Carlson 1990; Cooke 1990).

In the rural parts of the nation, the Census Bureau used USGS DLG-3 files. These cartographic files consist of data scanned from 1:100,000-scale maps (Callahan 1983; Sobel 1986). These maps, and thus the digital coordinate files produced from them, meet National Map Accuracy standards. Hays County, Texas was a county with TIGER/Line features rooted in 1:100,000-scale DLGs. Thus Hays County provides a useful laboratory to assess the quality of the USGS derived product.

In urban and rural areas, users have noted the following kinds of errors in the TIGER products. The Census Bureau compiled this list (LaMacchia 1990, p. 32):

1) New subdivisions were not on the maps;
2) Some streets on the maps did not exist on the ground;
3) The shapes of features were incorrect;
4) Some extraneous lines appeared on the maps;
5) Some map features were not named;
6) Some mapped features had the wrong names; and
7) Some features had the wrong symbology.

The seven errors fall into two distinct categories: cartographic and database. The cartographic errors (errors 1 through 4) require edits to the digital file containing the road coordinates. The remaining errors are attribute (e.g., street name, feature type) errors correctable in the database file holding the attributes of the roads. All three database errors, numbers 5, 6, and 7 can be made visible on the map (e.g., a map with all street names plotted), but are correctable on the attribute side. Error 2 has two forms that need to be treated differently. In its first form, Error 2 means that streets on the ground have been moved or removed because of road-alignments or route replacement. That correction needs to be made on the cartographic side. The second form of Error 2 applies to drives or unpaved tracks that may exist, but are private and thus should be reclassified. This form of Error 2 is correctable by changing attributes on the database side.

In the final TIGER/Line product, released after the 1990 census, all errors were to be reduced to no more than 5 percent (ibid., p. 32). What is the situation with regard to these errors in rural areas as represented by Hays County, Texas? Was the TIGER/Line file for Hays county 95 percent accurate? Which of the seven types of error identified by users occurred in the Hays County TIGER/Line file? Which of the errors were most common in the county? Does the TIGER/Line file for rural counties contain errors not on the Bureau list? This paper addresses these issues.

Methodology
TIGER/Line file accuracy and completeness was assessed from the cartographic and attribute perspectives described above. The assessment was made possible by comparing the TIGER/Line file maps with an independently generated digital map of Hays County produced between 1987 and 1990 (Rudnicki 1992). This database was compiled from 1:24,000-scale U.S.G.S. topographical quadrangle maps, dated between 1963 and 1986; tax appraiser aerial photos of the county taken in 1982; aerial photos of San Marcos, the largest city in the county, taken in 1985; and various subdivision plats filed between 1982 and...
1990. Extensive field work verified the existence of various roads and helped establish the horizontal control necessary to tie in new subdivision streets with existing roads. Political boundaries were compiled from the quadrangle maps and updated by information provided by local authorities. Roads by type, their names, and address ranges (where available), were the primary database attributes. The county's digital map was organized as a series of quadrangles or tiles each 7.5 minutes in latitude and longitude, reflecting the origins of the data set as topo sheets (Figure 1).

For the purposes of this paper, the assessment of TIGER/Line data was limited to road information. Hydrographic features were not examined. That component has fewer attributes and at the present time does not have the importance of road information to transportation planners and others. Hydrographic information is used for locational reference purposes only. The focus on street/road information corresponds with that expressed by TIGER/Line users (LaMacchia 1990).

Both the cartographic and database elements were compared. The comparison was meant to identify TIGER/Line file errors. On the cartographic side, maps of road arcs were generated from the Hays County database and the TIGER database. The two sets of maps were superimposed and compared. Recently built roads (to 1990) should be on the map (Error 1). Roads drawn on the map should be able to be traveled on (Error 2). In addition, roads were checked for correct shape (Error 3); and any spurious lines or points in the TIGER data set were noted (Error 4).

The comparison technique led to the identification of a new cartographic error, not previously identified by TIGER product users. This error involves positional accuracy: are streets situated in the correct location and in the proper orientation to other streets? This error is not apparent when looking at a TIGER/Line map by itself. When a second map of the same area and at the same scale is compared with the TIGER map or is superimposed on the TIGER product over a light table, the incorrect position errors become visible. These errors were most obvious on the map for the Buda area. In this paper, mispositioned roads will be identified as type 8 errors. This evaluation criterion is not available to users unless they have an independently generated map of the area being scrutinized.

The attributes investigated were those that related to the accuracy and completeness of the road features. Comparison of the TIGER data with the county's data consisted of two parts. First, the maps were compared. The TIGER product was scanned to find segments of road that had not been named (Error 5) and roads that had been misclassified and thus were symbolized incorrectly (Error 7). Second, lists of names were generated from the two databases and compared. Missing road names (Error 5) and misspelled names or misnamed roads (Error 6) were identified in this way. Missing names took two forms on the map. First, all arcs representing a road may be unnamed. In this case, the entire road does not appear on the map. In the second case, some arcs comprising a road would not be named. On the map, such unnamed street segments would show up as gaps along a road.

Attributes were not considered as extensively as the cartographic portion of the database, however. Attention focused on the attributes of census feature classification code (CFCC), source, feature name, and feature type. Many other attributes such as census tract number, block number, zip code area, congressional district, and address ranges were not examined.

An issue affecting cartographic and attribute elements is the date of the TIGER/Line information. The U.S. Geological Survey delivered their last digital line file to the Census Bureau in 1987, but the aerial photography used to update the DLGs was taken much earlier in the decade, 1981 and 1983 for Hays County. Considerable road building has occurred since that time not just in parts of the county but also in various areas across the United States. Thus, at best, the information from USGS was seven years old by the time of the census. Potentially, the amount of error in the TIGER/Line files for any given county could be considerable and would vary based on the amount of growth in the county and the date of the cartographic source materials. Error 1 would result from dated source materials.
A related potential error involves road classification. While the USGS did an impressive job compiling the digital cartographic database for the nation, they assigned road types based on aerial photo interpretation techniques. No field checking of the new data (1981, 1983) was done for Hays County. Thus road types in the digital database may be incorrect. Errors 2 and 7 would result from a lack of field checking. This situation would obtain in the rest of the rural United States as well.

The seven user-identified TIGER product errors have been classified as to whether they were cartographic or database. These errors can also be divided into two classes based on cost (time and effort) to correct. These two classes were labelled "high-cost" and "low-cost". High-cost errors require considerable time and energy to correct; low-cost errors do not. For example, high-cost errors require a modification of the cartographic database at a graphics work station by digitizing new streets or importing and integrating a cartographic data file produced elsewhere. Low-cost errors, on the other hand, could be corrected by clerical operations involving only an ordinary text terminal and the existing data files to correct the spelling of a street name or change the CFCC of a road arc. Errors 1 through 4 and the newly identified Error 8 are cartographic, and generally, high-cost. Errors 5 through 7, attributes, are low-cost. Depending on the type and extent of the errors, their existence can strain the resources of an organization in cleaning or data scrubbing operations that extend the amount of preparatory work before the data can be incorporated within a productive GIS framework.

Three representative areas of Hays County were selected for close scrutiny, each 7.5 minutes in extent. These areas were the Buda, Henly, and San Marcos North quadrangles (Figure 1). Buda represents a rapidly suburbanizing area, Henly represents a rural area with little change in

**FIGURE 1.**
Hays County, Texas. The database is organized on a tile basis with each tile corresponding to a 7.5' U.S.G.S. quadrangle. The Henly, Buda, and San Marcos North quadrangles represent three different line feature environments.
spatial features over time, while the San Marcos area (not dimensioned) represents a relatively built-up urban area. The software used for the analysis was pAcR-INF0. This software has the capability to extract quad-sized pieces from the TIGER county database. Using these smaller portions of the entire county database helped focus the analysis.

Analysis

Henzl (cartographic elements)

The Henzy quad was the most rural of the three areas sampled within Hays County. As might be expected, that part of the county had the fewest roads: 330 named road arcs on the TIGER/Line file map. While the named road arcs in the county's database are assigned in such a way as to define a continuous network, this was not the case with the named road segments in the TIGER/Line file (Figure 2a and 2b). The named road network appears disjointed. Roads appear to float in space, unconnected to other road segments. The non-visible road segments do exist in the database, but they have not been assigned names. Users of the TIGER/Line file must assign the appropriate names to complete the road network. This error is a low-cost one (Error 5). While this error manifests itself cartographically, corrections are made in the attribute portion of the TIGER/Line database.

The maps show that the census product (Figure 2b) contains a number of named, short roads that do not appear on the county map. The names given to these short roads, such as: Smith Ranch Road, McLaferty Ranch Road, and Freestone Ranch Road identify these as private drives leading to houses. Indeed, field checking showed that some of these roads are behind locked gates. Thus, these names are not the names of public thoroughfares, but rather names presumably assigned by census enumerators to help them find their way in the Texas Hill country. The source code assigned to these features was "E", Census Bureau Enumerator update. These road names are not officially recognized by the county, volunteer fire departments, or anyone else except the owner of the private drive. These misnamed roads are examples of Error 6.

A third error in the Henzy portion of the Hays County TIGER/Line file involves the shapes of several road arcs (Error 3). This problem is well illustrated by roads representing the Mystic Creek subdivision (Figure 3a and b). While two streets exist on the ground, the subdivision is portrayed as having only one street (unnamed) with a most unusual shape (Figure 3b). This shape is apparently due to the way the Census Bureau handled field updates. Census enumerators who found new roads added them to their field maps after estimating street lengths and directional orientations. These field drawings were then sent to Bureau regional field offices where the new roads were digitized into the TIGER/Line database. Thus, while new subdivision roads have been added to the post-census TIGER files, their shape cannot be assumed to be accurate. Similarly, the positional accuracy of the new roads cannot be assumed to be nearly as high as the original DLG data. Fortunately, these road arcs are identified by a source code of "E", Census Bureau Enumerator Update so they can be readily identified. In addition to the Mystic Creek subdivision road, 38 other arcs are tagged with that source code in the Henzy TIGER/Line file. Those road arcs should be checked for correct shape and positional accuracy (Errors 3 and 8).

It should be noted that this "error" was not perceived as such for Census purposes. For its operations the Bureau needed only to know that a street: existed and that it be located correctly relative to other streets. For its operations, the Bureau did not need to know the precise shape and exact position of a road relative to other roads. Potential users of TIGER/Line file data need to be aware of that orientation. If other users have the same orientation, then they too will not perceive shape and positional anomalies as errors.

Not all of the road arcs with a source code of "E" will have an incorrect shape or be located in the wrong place because as has already been noted, some enumerator updates consisted of only providing street names to arcs already present in the DLGs and nothing more. No simple way exists for differentiating among the different kinds of enumerator updates, so all road arcs with a source code of "E" need to be checked.
FIGURE 2a.
Named Road Segments.
Henly quad.
Hays County base map.

FIGURE 2b.
Named Road Segments.
Henly quad.
TIGER/Line File.
Correcting Error 3, incorrectly shaped road features, is a high-cost one compared to the previous two. Users who require an accurate representation of streets on their maps have to look outside TIGER to correct the difficulty. The subdivision’s streets need to be digitized and then merged with the TIGER cartographic elements. In addition, on the attribute side, the values for the various items, such as name and CFCC, need to be added for those new streets. In the Henly area, this cartographic difficulty potentially affects as many as 39 (i.e., up to 12 percent) of the 330 road arcs.

Finally, the TIGER/Line file map contained many more roads than the county map (Error 2). Attempts to find these roads in the field produced mainly negative results. Most of the TIGER/Line unnamed roads are behind locked gates and lead onto range land or to a hunting cabin. Some "roads" behind gates or fences were so overgrown that they were scarcely recognizable and apparently hadn’t been used in at least a year. These unnamed "roads" are private. These drives or tracks ostensibly are a low-cost error because only their CFCC attribute needs to be tagged as “private” (code A??) in the database portion of the TIGER/Line file, and so kept off the map until the roads’ status change some time in the future. Unfortunately, the Census Bureau has no classification code for "private drive" or "dirt track behind locked gate". A new road category needs to be devised by the Bureau for rural areas.

For the most part, the TIGER/Line file in the Henly area had low-cost errors which did not take long to "scrub." Field work, to verify which TIGER/Line road arcs to include in the database, took one person about one day. Clerical editing to modify attributes in the TIGER file database to provide a reasonable 1:100,000-scale product took about two hours. Correcting the shape of the Mystic Creek road, adding the missing road, and assigning attributes also took two hours. That time does not include the time and cost involved in acquiring orthophotography of the area, which was needed to correct the road existence and shape errors. Presumably, this type of data-scrubbing situation holds true across the rest of rural America as well.

**Buda (cartographic elements)**

The Buda, Texas 7.5’ quadrangle is much "busier" with 1281 road records as opposed to 330 records in the Henly area. This portion of the county has many fewer ranch roads that are so common in the Henly area because the Buda area is a deep-soiled, level to rolling farming area. Few roads exist in the plowed fields that characterize the area so the problem of

---

**FIGURE 3a & b.**

a) A portion of the Henly area as represented in the TIGER/Line file
b) The same portion as represented in the Hays County generated base map.
FIGURE 4a.
Named Road Segments.
Buda Quad—Portion.
Hays County base Map.

FIGURE 4b.
Named Road Segments.
Henly quad area.
TIGER/Line File.
named private drives observed in Henly is not a problem here. Similarly, no dirt trails behind locked gates exist here to be misclassified as public roads as was the case in Henly. Thus, the Buda area TIGER/Line file contains a fairly complete set of "legitimate" road arcs when compared with the county map. While no road arcs have to be deleted (Error 2), a number of them need to be added (Error 1).

This part of Hays County lies just south of the Austin metropolitan area. Proximity to a large city and relatively flat land helped the Buda quadrangle area grow rapidly, particularly in the early 1980s, as Austin's suburbs grew. The most recently published USGS Buda quadrangle dates from 1973, so this area is a good one to examine for missing subdivisions. Few were found. In fact, the TIGER/Line file contains many of the recent subdivisions and new streets.

The Buda area TIGER/Line file contains five cartographic errors. Four of these errors can be assigned to categories identified by TIGER/Line file users. The fifth error involves the positional accuracy of new subdivision streets. The four errors from the Census Bureau list include: missing subdivision streets (Error 1), floating road arcs (Error 5), inappropriate cul-de-sac symbols (Error 3), and inadequate line splicing (also Error 3). The fifth error becomes apparent when the two maps produced from different sources are examined side by side (Error 8).

Most of the Buda quad is shown in Figure 4. Zooming in slightly has clipped several roads along the map's margins. The enlargement permits focusing on two subdivisions, one in the northwest and the other in the southeast portion of the quad (Figure 4a and b). Comparison of the two maps shows that the TIGER/Line map is missing a number of subdivision streets. In the northwest, the Coves of Cimarron subdivision is inadequately represented, and in the southeast, streets are missing from Goforth Village subdivision (Error 1). These errors are high-cost ones because solutions are expensive, requiring the acquisition of an ortho-photo and digitizing from it. This type of error is more serious in the Buda area than Henly because more subdivision building activity occurred around Buda during the 1980s.

The second error visible on the BUDA TIGER/Line file map is the floating street phenomenon previously encountered in the Henly area (Error 5). This is a low-cost error because the digitized road arcs exist in the TIGER/Line file. The street segments simply need to be named. Frontage roads along IH-35 are examples.

Two other distinctive errors on the Buda map, not present in the Henly area, include mapping cul-de-sacs as small circles (Figure 4b), and a misalignment of adjoining road arcs (Figure 5). Both of these cartographic inaccuracies have been classified as type 3 errors and are a result of patching together independently generated pieces of the map database by different people at different times under differing assumptions. The TIGER/Line file is a collection of five different digital map products.

1) DIME files from the 1970s,
2) 1:100,000-scale digital line graphs (DLG) produced by the USGS in the 1980s,
3) Census Bureau updates in various regional offices during the mid to late 1980s,
4) Contractor digitized 1:24,000-scale "connections" between the urban DIME and rural DLG sets, and
5) Census Bureau enumerators making field measurements of new subdivision streets that weren't captured in products 2, 3, or 4 (Guptil 1984; LaMacchia 1983).

Overall, products 1 and 2 are the basis for the bulk of the TIGER/Line files, as has already been mentioned. Products 3, 4, and 5 play supporting roles.

The cul-de-sac circle symbols appear to be generated by contractors (#4) or Census enumerators (#5). The cul-de-sac circle symbol may have been acceptable at the digitizing scale of 1:24,000, but at a scale of 1:100,000 the circles are reduced to small blobs. Examples are found south of Buda (Siesta Verde subdivision) and in the Woods of Cimarron, just north of the Coves of Cimarron (Figure 4b). These are low-cost errors because graphic workstation operators can easily reshape the circles into short line segments (Error 3).

The interrupted road line problem is illustrated in Figure 5. The interrupted road lines mark the inadequate splicing of products 2 and 4. The TIGER/Line file contains separated sections of the same road. The split occurs across the middle of the map, so that roads on the top
half of the TIGER map are shifted slightly to the left of the roads on the bottom half of the map. Three line segments representing IH-35 and its frontage roads on the left, and County Road 119 on the right are shown in Figure 5. The reason for this break becomes apparent when the sources of the segments are examined. The road arcs on the top half of the map are identified as having the 1980 GBF/DIME file as their source, while the road arcs on the bottom half of the map have “USGS 1:100,000 map” as their source. This information is surprising because the extended DIME portion of the TIGER/Line file, as the contractor produced 1:24,000-scale “connection” was called, was supposed to have merged with the DLG data at a county line or at the boundary between quad sheets (Adamik 1990). This is a low-cost error and local users of the TIGER/Line file can edgematch the misaligned line segments fairly easily at a graphics workstation.

Finally, for local users to whom accurately located roads are important, 15 percent of the 1281 road arcs in the Buda quad area have a source code of “E”, Census Bureau Enumerator Update. As in the Henly area, these road arcs need to be examined closely to ascertain their true shape and ground positions (Errors 3 and 8). Some position errors are obvious even at the small scale of the accompanying map, particularly in subdivisions (Figure 4). Several shape errors are visible as well, but their deviation is modest compared with the error found in the Mystic Creek subdivision in the Henly area (Figure 3). These high-cost errors will take a considerable amount of time to correct by local agencies wishing to use the TIGER/Line files.

San Marcos North
(cartographic elements)
San Marcos, the largest city in Hays County with a population of 28,743, straddles two 7.5' quadrange maps, the San Marcos South and the San Marcos North sheets. For the purposes of this sampling of the TIGER/Line file, the San Marcos North quad was selected because it contains most of the city’s streets. The quad is thus representative of the urban part of Hays County. Examining the San Marcos South quadrangle as well as the San Marcos North quadrangle would have added little new information.

As might be expected, the San Marcos North area has a denser road network than the Buda area, with 2762 road segment records. This represents a 116 percent increase over the number of road segments in the Buda area. Yet enumerator updates (Source code of “E”), the cause of shape and positional errors, were relatively fewer in this portion of the county. In the San Marcos North area, 7 percent of the road arcs have the enumerator update code.
versus 15 percent in the Buda area. That 7 percent translates to 203 road arcs that should be examined for positional accuracy and shape distortions by users who seek an accurate database. The reason for the lower percentages is that the subdivisions of the San Marcos North quad tend to be older than those in the Buda area and hence are well represented on existing maps. Most of the streets within the city limits of San Marcos have also existed for a number of years. Those streets were named and paved years ago.

Consequently, street maps with names for most streets are readily available and were probably used by Census staff at the time of DLG conversion into TIGER/Line format during the mid-1980s. Because the map base was in better shape when it reached Census enumerators, they had relatively fewer corrections and edits to make.

As in the case of Buda, the San Marcos maps represent most but not all of the quad so that certain TIGER/Line map features can be emphasized (Figure 6a and 6b). Three categories of error are significant in the TIGER/Line rendition of the San Marcos North quad. These include: floating roads/missing roads (Error 5), non-existent streets (Error 2), and features with incorrect shapes (Error 3).

Error 5 manifests itself in a striking manner as a large area devoid of roads in the northwest part of the TIGER map relative to the Hays County map. While highly obvious visually, this gap represents a low-cost error. Clerical editing of the database can easily change those non-visible arcs into mappable ones. Clerical editing will also make the IH-35 frontage roads visible, as well as some streets on the south side of the small town of Kyle. Clerical editing will anchor several “floating” roads as well. The TIGER/Line file as database for San Marcos North appears reasonably complete, as far as the cartographic rendering of roads are concerned.

TIGER/Line file users need to be careful because the file contains a number of the unnamed road segments do not exist on the ground (Error 2). As was the case in the Henly area, careful field checking is required to identify which roads are public and which are private. The private roads need to be tagged appropriately so they won’t show up on maps generated from the TIGER/Line file.

Here in the urban portion of the county, users will be most tempted to increase the scale of the TIGER/Line file line map far beyond what its developers intended. To make a city map and utilize the names that exist in the TIGER database requires a large scale. It is at this large scale that the final error type (Error 3) becomes most apparent (Figure 7). This “notorious Y” problem is particularly vexing for streets that lie on the diagonal rather than the vertical or horizontal. This dislocation of intersections is a function of the raster-to-vector conversion process used in the production of DLGs by the U.S. Geological Survey (Domaratz 1991). On the surface the “Y’s” appear to be a low-cost problem. In practice, the manual correction of each intersection is extremely tedious.

Road Attribute Data Quality

User-identified TIGER product attribute errors included unnamed features (Error 5), incorrect names for mapped features (Error 6), incorrect feature classifications (Error 2), and incorrect feature symbology (Error 7). Error 7 was not found in the portions of the Hays County TIGER/Line file sampled in this study; i.e., road arcs are plotted as road arcs and stream arcs are plotted as stream arcs. Errors 2 and 5 have already been dealt with, because they manifest themselves in map form. Thus, the following discussion deals with error 6. To assess road/ street name errors, two lists of road names for the three areas were generated, one from the TIGER/Line file and the other from the Hays County database. These names were then compared and differences and omissions noted.

In Henly, each of the two databases contained 30 unique road names. Of those, 16 were in the TIGER database but not in the Hays County database; while 18 names were in the Hays database but not in TIGER. Thus about 50 percent of the road names need to be modified in TIGER. Some road names were counted as incorrect because although they appear in both files, the names may be spelled differently or the road type may differ, for example, Smith Drive instead of Smith Avenue.

42 URISA journal/Refereed
FIGURE 6a.
Named Road Segments.
San Marcos North quad—portion.
Hays County base map.

FIGURE 6b.
Named Road Segments.
San Marcos North quad—portion.
Tiger/Line file.
The Henly TIGER street names are interesting because, as has been mentioned earlier, the Census enumerator named a number of private drives for their ranch owners. The TIGER street map contains Freestone Ranch Road, McCall Lane, P. Fullman Ranch Road, Smith Ranch Road, and Josal Drive. These names are not the names of public thoroughfares. The Henly area, one of the most rural parts of Hays County, has proportionately the highest street name error rate. This part of the county is not an established area like San Marcos. A certain amount of independent street naming was not untoward by Census enumerators who were trying to make sense of a chaotic settlement pattern.

The approximately 15 different road names suggesting a 50 percent error rate in the Census TIGER/Line file is perhaps overstated for another reason. For example, eight of the roads are identified in the TIGER database by their county road designation, e.g., CR191. In the Hays County database the road’s name is given as Gatlin Creek Road, with CR191 listed as an alias. If those eight roads are interpreted as having been named correctly, then the error rate in road naming in TIGER drops from 50 to 26 percent. The latter figure is still above the 5 percent margin of error that the Bureau expected to see in the Version 3 (post-Census) TIGER. Misspellings appear to be absent although for some named roads the only way to know for sure would be to ask the ranchers for whom the road was named.

The Henly TIGER street name file illustrates another common problem in the TIGER attribute file, inconsistent naming. For example, some road arcs are named “Hays Co Rd 191”, while nearby arcs are named “County road 191”. While not a major difficulty, agencies will have to devote clerical resources to cope with this low-cost error which is peppered throughout the TIGER database.

The Buda quad contained approximately 200 named roads (191 in TIGER and 235 in the Hays database). Of those, 47 were found in the TIGER database but not in the county database, while 92 were found only in the county database. Thus approximately 32 percent of the road names in TIGER were incorrect.

These mismatches consisted of streets that did not exist in the TIGER database, street names that were similar in spelling in the two databases but not exact, and streets that were identified as “drives” or “roads” or “avenues” in the Hays County database but were listed as different street types in the TIGER database. These errors are of the low-cost variety because they can be corrected by simple clerical operations.
The Buda area has a higher names error rate than either the Henly or San Marcos North areas partly because this area had considerable subdivision activity during the 1980s. New roads and their names were not captured in the TIGER database.

The San Marcos North quad had some 375 street names. Of these, 41 were unique to TIGER, and 79 were found only in the Hays County database. Thus about 13 percent of the names need to be corrected by either adding a road name to an unnamed street segment, correcting the spelling of a road name, or changing the road type (street, drive, road, ...) in the TIGER/Line file. Proportionately, this area, with the largest number of street names, had the lowest street name error rate. This was expected because, as already mentioned, the San Marcos North quad area has been settled for a long time, most roads are paved, numerous street signs stand at street intersections, and maps with street names exist of the area.

Summary and Conclusion

Six different TIGER/Line file errors were identified in this study. Five had been noted previously by TIGER/Line file users. Some may judge one user-defined error found in Hays County not an error, i.e., one could argue that all streets on the TIGER/Line map exist on the ground; but whether some of these are "streets" is arguable because they turned out to be private drives or dirt tracks, often behind locked gates. When the TIGER/Line file map was compared with an independently generated map of Hays County, one additional error became apparent: the incorrect position of some roads on the ground. This error was confined to new roads and streets that added to the TIGER database by Census enumerators using crude surveying techniques.

The Census Bureau's goals do not appear to have been reached. The Bureau had hoped for a 95 percent accurate data product. The cartographic product does not approach 1:100,000 map accuracy standards for 7 to 15 percent of the road arcs in Hays County. The attribute product contains between 13 and 50 percent errors. The Bureau's predicted accuracy rate appears to have been over-ambitious given the short time for preparation of this national database. On the attribute side, although road name errors ranged from 13 (San Marcos North) to 50 percent (Henly), these errors are low-cost. The attribute problems are minor given the magnitude of this digital mapping project. The fact that so many roads are named, and for the most part named correctly, is a significant accomplishment. Adding road names is one of the major contributions of the Census enumerator. In the precensus TIGER/Line file for Hays County very few roads were named. In the post-census file these roads are named and are listed as having "Census Bureau Enumerator Update" as their source.

On the cartographic side, the error rate is lower but more difficult to correct. Three shortcomings of the cartographic portion of TIGER include positional inaccuracies, distorted road shapes of roads, and the incompleteness of the data set due to continual modification of the landscape by road building, road realignment, and road abandonment. To make a street map of the nation that is current when it reaches users seems an impossible task using existing map-making technology. The situation calls for local cooperation, yet how many rural local governments have a GIS, one built on an accurate and current cartographic base?

Users are generating creative solutions to the cartographic problems in the TIGER/Line files. For example, Global Positioning System (GPS) technology is being employed to add roads missing from TIGER (Lange and Kruvenski 1990). The same technology could correct the shapes of roads and perhaps identify and correct road positional inaccuracies. Artificial intelligence approaches are being evaluated to identify road arcs that have an inappropriate shape (Wong 1991). No way to identify positional inaccuracies exists, other than to compare the TIGER map against an existing map. Such maps are available in the form of SPOT satellite panchromatic images. Using sophisticated software, the raster format data from the image can be used to edit the vector format TIGER/Line data (Bauer 1992).

What makes some of the cartographic errors so noticeable is the abundance of the attribute information associated with each of the cartographic ele-
ments. If users were content to manipulate only the attribute portion of the TIGER/Line file database, there would be fewer complaints; but where the data are being manipulated in graphic form, the less than pleasant appearance of the TIGER map is disquieting to people accustomed to working with maps (Cooke and Levasseur 1990). Unfortunately, larger scale maps are very likely when using TIGER data because of the large amount of road attribute information. For example, road names and address ranges cannot be plotted at the small scale of the original cartographic database. The feature names need to be plotted on a cartographic base that approaches and even exceeds 1:24,000.

A larger scale map exacerbates the cartographic shortcomings stemming from the relatively low resolution of the TIGER/Line files: 1:100,000-scale. Maps from TIGER at a larger scale also makes visually disquieting the inadequate digitizing/scanning technology that caused intersections to be dislocated ("Y" rather than "T" shaped).

The TIGER/Line file database is thus unbalanced in a major way. The cartographic side is lacking while the attribute side is rich. For applications that depend on addressing, routing, or allocation along a network the TIGER/Files offer a growth path. No addresses exist in the records of the TIGER database for Hays County. The data fields are in place, however. Local agencies need only assign addresses to rural areas currently located by route and box numbers, enter these addresses into the TIGER database, and database querying and analysis can proceed, in spite of a cartographic rendering that is less than elegant.

The Census Bureau is aware of the need to improve the TIGER database. The Bureau has initiated discussion of a Proposal for TIGER Data Exchange program (Bureau of the Census 1991). The goal is to explore the feasibility of integrating local area corrections and updates with the national TIGER database.

Never before have spatial data processors had such an information-laden cartographic database for the entire nation. The product was released before it was ready; it is, in effect, a first draft. The product has shown improvement between releases, with the latest release (version 3) being the most complete. The database has potential for rural counties not having access to any other database. These counties and small cities, with some time and effort, can adapt the TIGER database to their geographic information processing needs.

Acknowledgements

I'm indebted to Robert Burke who played a major role in the data analysis of the pre-census TIGER files for Hays County, Texas. Chitra Subramaniam-Bryson identified inconsistencies between the Hays County DLGs and TIGER/Line data. Michael Garza evaluated TIGER/Line file accuracy using GPS-derived positions. Mohammad Tangestani generated most of the maps.

References


Integrating Dynamic Modeling and Geographic Information Systems

N.W.J. Hazelton, F.J. Leahy and I.P. Williamson

Bill Hazelton has recently completed a Ph.D in Temporal (4-D) GIS in the Department of Surveying and Land Information, University of Melbourne, Australia. He holds a B.Surv. from the same department and has worked in surveying and mapping. He has worked in Antarctica as a glaciologist, travelling extensively on over-snow traverses within the Australian Antarctic Territory.

Frank Leahy is a senior lecturer in the Department of Surveying and Land Information, University of Melbourne. He has published and consulted extensively in geodetic surveying, mapping, LIS/GIS, satellite geodesy and related areas.

Professor Ian Williamson holds the chair in the Department of Surveying and Land Information at University of Melbourne. He has published and consulted extensively in the LIS/GIS area.

All three authors are licensed surveyors and members of the Institution of Surveyors, Australia.

Abstract: A major limitation to the development of GIS is that most systems model reality in the same way that a paper map models reality. Until GIS can develop into direct ‘reality modelling’ systems, rather than ‘map modelling’ systems, they will not be able to handle 3-D spatial data, nor temporally-referenced data.

A GIS is generally used for some level of decision support. Decision-making relies on the analysis of various options, which are based upon the likely outcomes of different management strategies in the enterprise under consideration. Some form of dynamic modeling is required to simulate future scenarios.

Dynamic modeling cannot be linked with a GIS at present because of incompatibilities in the conceptual frameworks of the two systems. A multitemporal 4-D GIS would allow the integration of dynamic modeling and a GIS. In this paper, it is argued that the development of a true 4-D GIS is one of the “Grand Challenges” facing GIS in the near future.

In late 1991, David Mark of the State University of New York at Buffalo (an NCGIA Centre) asked whether there were any “Grand Challenges in GIS.” Subsequently, the authors of this paper have identified three GIS challenges:

1) That GIS must develop beyond a method for representing maps in a computer, toward a means of representing reality directly as it can be perceived by humans, rather than through an intervening medium (such as a map representation);
2) That as a consequence of (1), GIS must become able to handle 3-D spatially referenced data, temporally referenced data, multiple interpretations of the same reality, and different views of reality dependent upon time;
3) That also as a consequence of (1), GIS must be able to provide decision-support services by such means as predictive and simulation modeling, spatio-temporal analysis, and support for multiple simulated futures.

These three challenges will not be everyone’s choice as there are many worthy “Grand Challenges” for GIS. In this paper, it is argued that at least one of them is to develop a four-dimensional GIS. This paper will discuss why a 4-D GIS is a necessary development for current GIS.
The third challenge mentions GIS in providing decision-support services. This application of GIS focuses attention on the roles of modeling and decision-making in GIS use. Cowan’s (1988, p. 1554) definition of GIS as a “decision-support system involving the integration of spatially referenced data in a problem-solving environment” emphasizes the use and purpose of GIS, rather than just the technology used to build a system—what GIS should be, rather than the nuts and bolts of its construction.

In this paper, we will consider that a ‘whole’ GIS consists of five components: hardware, software, data, people, and procedures (Dangermond 1988; Turk 1988, 1990, 1992).

**Decision Support and GIS**

What is involved in making decisions? The essence of decision-making seems to lie in making informed choices among a number of options. ‘Hobson’s choice’ requires no decision-making. Decision-making may be improved by increasing the number of options in the hope of getting better ones, improving the information about the choices, or improving the mental processes involved in making the choice.

Increasing the number of options involves being able to model the consequences of as many decisions as possible; this is a task for some type of predictive model. Improving the information about the various choices involves analysis of the options and comparison between them; this is a task for analytical tools. Improving the mental process is beyond the scope of this paper. Turk (1990, 1992) discusses aspects of this issue.

An example of a simple computer-based decision-support system is a spreadsheet. The user can alter many variables in the models of future options, compare and analyze the results of these models. The user can derive a larger number of models and better analyze them, than by using manual techniques. In analyzing the state of a business enterprise, for example, many different models of interest, tax and depreciation rates may be used to produce a broad selection of options for decision-making.

At the simplest level, a spatial decision-support system (SDSS) is expected to generate and analyze multiple outcomes of possible management options. This requires some form of predictive model, a model in which time is a parameter; such a model is referred to as a dynamic model.

A SDSS requires not only a GIS to manage the data, but the ability to operate with time as a parameter, handle dynamic models and their output, and manage spatio-temporal analysis of data. This requires integrating dynamic models with GIS.

**Modeling and GIS**

In the broadest sense, modeling is an integral part of GIS, but often seems to be treated as though it were quite a separate process. A GIS attempts to create a model of the ‘real world’ or, at least, certain aspects of it. In this section, a brief overview of current GIS modeling concepts is presented in order to provide an understanding of how GIS and modeling fit together.

Modeling is a very general term. In its most general sense, modeling is the creation of an abstract process that attempts to emulate another, often more complex, process. The mathematical model “1 + 1 = 2” is an abstraction process that describes how people can count two of any type of object. The mathematician Hamming has stated that mathematics is a remarkably good abstraction of reality. He notes (1980, p. 81) that:

I have tried, with little success, to get some of my friends to understand my amazement that the abstraction of integers for counting is both possible and useful. Is it not remarkable that 6 sheep plus 7 sheep make 13 sheep; that 6 stones plus 7 stones make 13 stones? Is it not a miracle that the universe is so constructed that such a simple abstraction as a number is possible? To me this is one of the strongest arguments of the unreasonable effectiveness of mathematics. Indeed, I find it both strange and unexplainable.

Models have several constraints, among them: they should be operationally simpler than the process that they emulate; they should be quicker than the original process in most cases; and you should be able to use them without affecting the state of the original process.

Modeling is a very broad concept and people have differ-
ent ideas of what is meant by the term, so one can begin by categorizing modeling. These categories are by no means exhaustive, nor are they mutually exclusive. However, they include most GIS modeling concerns and allow a look at some general characteristics that are associated with specific types of models.

Within a ‘whole’ GIS, modeling may be generalized into four categories:

1) **Reality Modeling:** The means by which a GIS stores a model of ‘reality’. Different GIS use different representations of ‘reality’, e.g., grid-cell structures, discrete irregular polygons, and 2-D and ‘2.5-D’ representations of the surface of the Earth. This is primarily a function of GIS software.

2) **System Modeling:** How a GIS as a whole will work, its relationship to the rest of an organization, data requirements, outputs, efficiency.

3) **Empirical Modeling:** Most current operations in GIS, especially queries, proceed using an empirical model.

4) **Dynamic Modeling:** These models utilize time, as well as some form of deterministic, stochastic or chaotic model, to generate future scenarios.

**Reality Modeling**

Reality modeling is the model that the GIS ‘has’ (or was designed with) of the ‘real world’. Unfortunately, current GIS tend to be map-representation systems, rather than reality-representation systems. This can severely limit their view of the world and the types of reality models that are possible within GIS. Many of the reality models in current GIS are very close to paper maps.

**System Modeling**

This type is concerned with modeling the GIS, rather than working with data within the GIS. The types of things modeled include: the operations that the system will be required to perform; the data required for various operations; the usage patterns of the hardware and software; the communications within the ‘whole’ GIS (e.g., how the GIS meshes in with the rest of the organization); and the types of GIS specialists that are required to operate the system. A discussion of some aspects of LIS/GIS system modeling can be found in Williamson and Hunter (1991).

As will be realized, most of these models will tend to be conceptual models of the system and will, or should, be undertaken before the purchase of a GIS is considered. As this is not the main area of interest in this paper, we will only note its critical importance to the overall efficiency and effectiveness of a ‘whole’ GIS.

**Empirical Modeling**

‘Empirical’ means resting upon trial or experiment, or known only by experience. In the GIS case, empirical modeling is that derived from experience of the real world, rather than from purely theoretical considerations or a priori reasoning. Burrough et al. (1988) divide modeling into the two general categories of ‘empirical’ and ‘process’ (called ‘dynamic’ in this paper).

Burrough et al. (1988) further divide empirical modeling into two general sub-categories. These are threshold models and regression models. There are other possible classification systems, but this is a more useful taxonomy for GIS.

Threshold models use the boundary values of specific attributes in the GIS to determine output values. The actual values of these boundary points are determined by experience. For example, it may be decided from experience that land suitable for residential development has the following attributes: slope < 7 percent; soil fertility less than 6 (on some previously decided scale); is not currently designated as prone to flooding; is within 500 meters of a major road; is within 800 meters of a public transport connection; and is within 1,000 meters of water, gas, electricity and sewerage connection points. These threshold values are then used to develop a model to determine suitable land from the data in the GIS. One may go further and decide to provide ‘gradations’ of suitability for each attribute. One may employ ‘fuzzy’ set theory to reduce the loss of information that tends to occur with the strict application of purely Boolean algebra.

Regression models use the results of experimentally observed regression analyses to decide the control parameter values. In this case, the resulting values are some function of the weighted values of the input values. For example, a regres-
sion model for residential suit-
ability of land may be:

\[ S = \frac{\text{Flooding}}{2 \times \text{slope} + 3 \times \text{fertility} + 3 \times D_{\text{road}} + 5 \times D_{\text{pubtrans}} + 4 \times D_{\text{services}}} \]

This is in very approximate correspondence to the example given for threshold models. GIS are well suited to the use of this type of model, as well as being able to generate the data for the initial regression analysis.

Empirical models comprise practically all models used with current GIS and, effectively, queries proceed based on some form of empirical model. Further, current GIS are designed to support this type of modeling at a very general level.

**Dynamic Modeling**

Dynamic models include various types of models referred to as process models, simulations, predictive models, etc. These models generate some form of prediction of the future behaviour of a system. The prediction may be a series of future states, the expected probability of future states of the system, or some other descriptive forms. In most cases, dynamic modeling uses time as one of its arguments.

Dynamic models for GIS would tend to generate future scenarios from existing ones, and a set of rules for prediction, which determine the type of dynamic model used. Burrough et al. (1988) describe two general types of dynamic models: ‘deterministic’ and ‘stochastic’. Hazelton (1991b) adds ‘chaotic models’ to these.

A deterministic dynamic model represents some observed relationship in terms of various processes in the system under consideration. The resulting values are generally determined by solving differential equations describing the process. The basic concept of a deterministic model is that it attempts to model the real world from fundamental physical or chemical laws, i.e., from some theoretical basis, rather than from an empirical basis. Deterministic models often build up a model of a system from many smaller models, rather like stones in a dry stone wall.

A stochastic dynamic model is used to model the average behaviour of a large number of events, based on probability theory. If a stochastic process, such as a Markov chain, is used to model the system’s behaviour, then the model predicts not an exact result, but one probable realization of many possible outcomes. The model also should provide some statistical estimate of the actual likelihood of this specific outcome.

The traditional examples of the different behaviours of a physical system are the clock and the coin-flip. “Utter regularity and utter randomness are the dynamical legacy of two millenia of physical thought.” (Crutchfield and Young 1990). However, time-dependent behaviour necessarily incorporates elements of both these processes. In ergodic theory, the formal model of a random process is the Bernoulli flow, \( B \). A completely determined Laplacian system, such as a perfect clock, can be denoted as \( P \), where the completely predictable pattern is repeated every \( t \) seconds.

\( B \) and \( P \) can be considered as the basic processes with which to model the complexity of non-linear systems. If some system is found to exhibit some repeating characteristics in the observed data, the observations can be described as having been produced by some variant of \( P \). If they are completely unpredictable, their generating process is essentially the same as \( B \). Any real system \( S \), will contain elements of both processes. We can ask, “Is it always true that the case of some observed behaviour can be decomposed into these two separate components?” Is \( S = B \otimes P \)?

Ergodic and probability theories state that in general this cannot be done so simply. It has been shown that “there are ergodic systems that cannot be separated into completely random and completely predictable processes (Ornstein 1989). The Wold-Kolmogorov spectral decomposition states that although the frequency spectrum of a stationary process consists of a singular spectral component associated with periodic and almost periodic behaviour and a broadband continuous component associated with an absolutely continuous measure, there remains other statistical elements beyond these (Kolmogorov 1977a, 1977b; Wold 1954; Crutchfield and Young 1990, p. 224). These statistical elements are generally described as ‘chaotic’.

As with building a dry stone wall, finding a suitable model for part of a larger modeling system is a matter of searching the pile of available stones until
a stone of the appropriate size and shape is found. Using the wrong stone distorts and weakens the wall and may cause it to collapse. Bulldozing stones together does not form an effective wall, nor does careless placement of the stones. Model building, like building stone walls, is as much art as science; an acquired skill that requires patience and knowledge.

Dynamic modeling is a rather more uncertain area than empirical modeling, and should not be approached casually. The need for much more data than are required for empirical modeling places pressure on the most expensive area of GIS operation—data collection—and introduces the need to consider surrogate data sets.

In summary, thus far:

- Modeling is the process of producing an abstraction of the ‘real world’ so that some part of it can be more easily handled. Most models are designed to represent a simplified version of some natural process or situation. Modeling and GIS are inseparable, GIS itself being a means of modeling reality. Unfortunately, most current GIS model paper maps, rather than directly modeling reality.

- Modeling is an intrinsic part of GIS, both within the ‘software’ component, and as a part of the ‘procedures’ component of a ‘whole’ GIS. System modeling is an important part of the conceptual development of a ‘whole’ GIS, and is used to help design the GIS.

- Empirical modeling is the imposition of certain empirical models from the ‘procedures’ component of a ‘whole’ GIS upon the ‘data’ component; this is used to derive new information from existing data. All current GIS allow this type of modeling, and it is the basis of most queries of current GIS.

- Dynamic modeling is the extension of empirical modeling to include the time domain. In this type of modeling, time is an essential parameter of the model. Few current GIS can handle temporally referenced data in any quantity.

**Dynamic Modeling**

Dynamic models are a subset of mathematical models, which are in turn a subset of symbolic and procedural models. ‘Dynamic models’ can be used as a generic term to describe all models that use time as a parameter. Table 1 shows the major classifications of mathematical models.

The most important classifications of models can be considered as contrasts between what are almost diametrically opposed design factors: distributed vs. lumped, dynamic vs. static, continuous vs. discrete, and deterministic vs. stochastic vs. chaotic. These factors affect the manner in which the model is designed and built, the types of problems that it can handle and its implementation on a computer. It also influences how the model can be tied into a GIS.

Of particular interest to GIS are distributed models, which use spatial location as a parameter in the model, although the discrete nature of data representation in a GIS means that truly distributed models cannot directly use GIS data. GIS can be used to prepare data for use by lumped parameter models, although this, in some senses, degrades the use of the GIS. Current GISs are probably best suited to ‘distributed-lumped’ models that allow lumped models to be used for small regions over the entire GIS data structure. Table 2 shows the internal requirement for lumped and distributed models.

The ‘distributed-lumped’ modeling approach is sound, provided that the models are supplied with data that describe the spatial variation of the GIS attribute data as accurately as the model can describe the process it is modeling. If this is not so, then the model’s results will be unreliable. The costs and benefits of quantified resource assessment and improved decision support need to be considered carefully, to ascertain

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Major Classification Divisions of Mathematical Models (from Kecman, 1988, p. 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Classification</td>
<td>According to</td>
</tr>
<tr>
<td>Linear</td>
<td>Non-linear</td>
</tr>
<tr>
<td>Time invariant</td>
<td>Time variant</td>
</tr>
<tr>
<td>Lump</td>
<td>Distributed</td>
</tr>
<tr>
<td>Flow</td>
<td>Batch</td>
</tr>
<tr>
<td>Deterministic</td>
<td>Stochastic</td>
</tr>
<tr>
<td>Static</td>
<td>Dynamic</td>
</tr>
</tbody>
</table>
TABLE 2.
Internal Requirements for Lumped and Distributed Models. The horizontal 'axis' of the table can be considered to indicate the degree of spatial representation in the model, while the vertical 'axis' of the table represents the degree of temporal representation in the model. (ODE = Ordinary Differential Equations; PDE = Partial Differential Equations). (From Kecman 1988, p. 12).

<table>
<thead>
<tr>
<th>Lumped Parameters</th>
<th>Distributed Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(x, y, z) )</td>
<td>( f(x, y, z) )</td>
</tr>
</tbody>
</table>

- algebraic equation representations:
  - tables
  - nomograms
  - characteristics

- Static
  - system of 1st order ODEs
  - ODE of n-th order
  - several ODEs of varying order

- Dynamic
  - one-dimensional ODEs
  - multidimensional elliptic PDEs

  - hyperbolic PDEs
  - parabolic PDEs
  \( f(t) \)

where and when the additional investment in data and computing resources will be justified.

Dynamic models can be deterministic, stochastic or chaotic in nature. Deterministic models are designed to produce 'exact' results, while stochastic models have their parameters determined by probability distributions. Chaotic models arise when situations of considerable complexity are modeled. The chaotic factor is not designed into the model, but appears by itself as a function of model complexity.

The complexity of a model is a function of the complexity of the system that it is trying to model. Because highly complex systems are not isomorphic to a purely deterministic model, a purely stochastic model, or any combination of the two (Crutchfield and Young 1990), it is necessary to consider the additional effects on the model caused by the complexity of the system. Complexity theory can be used to model some of these effects, and much of this modeling is complementary to the modeling of these effects using information theory.

Why Dynamic Modeling is Needed in GIS

"The amount of time spent by generalists in making technically based decisions is in inverse proportion to the complexity of the subject matter." Jones' Seventh Law, (Jones 1990, p. 176).

The quotation illustrates the problems of decision-making in an increasingly complex society. Specialists gather data in their respective fields, but because of the depth of knowledge required in many specialist areas, they often do not have a broad view beyond their own discipline. Generalists, on the other hand, do not have the depth of knowledge of the specialists, but strive to obtain an understanding of the whole situation. This is becoming increasingly difficult. As the quotation suggests, if the matter is highly technical, the generalists spend little time on the technical side, effectively leaving the decision to the specialists. This situation is caused by poor, or non-existent, decision support and communication between specialists and generalists.

Rhind (1988) has pointed out, "virtually all GIS developments thus far have resulted in 'data retrieval and sifting' engines; modeling work has not yet been brought together with this technically accomplished sub-culture". Densham and Goodchild (1989) have discussed the use of GIS as a decision-support tool. They note that current GIS have limited analytical capabilities and do not support decision-making adequately. Densham and Goodchild (1989) suggest that important areas of consideration for a GIS-based decision-support system include analytical modeling and expert knowledge.

Goodchild (1987) has suggested that the ability of GIS to adopt an extensive set of spatial analysis tools is limited by the inadequacies of the data structure, and the potential for development in output from GIS is limited by the query modes available to GIS, which are in turn limited by the cartographic model underlying the design of GIS in general. Goodchild (1991) also discussed the means of linking GIS and spatial data analysis concluding that the diversity of GIS data structures and the "unbounded nature of spatial data analysis" meant that development of a common
interface was unlikely. However, he felt that a close coupling of GIS and spatial data analysis was the best method of improving the current low level of GIS functionality in spatial analysis, together with making the discrete nature of GIS data structures explicit in the development of spatial analysis tools.

It has been recognized that a major intention of GIS is to act as a decision-support tool (Cowan 1988). The degree of support provided is often a matter of debate. Decision support can range from the provision of a topographic map, to a screen display of a modeling and analysis outcome that says "Do This." 'Advanced' decision support requires that the analysis of information in a GIS be a component part of a larger information system (the 'procedures' component of a 'whole' GIS). Hopper (1991) has pointed out that there is a dearth of study into the total flow of information through organizations, and of research into the value of information per se. A significant part of decision support is attempting to evaluate the consequences of management decisions or policies; this requires some form of predictive modeling.

Decision support means that all sides of an argument, all reasonable outcomes of a management decision, must be considered. The 'technological determinism' of Veblen (1924), where basic decisions are shaped by available technological capacity rather than the traditional political process based on ideology and value systems, has led us to a polluted, un-happy world, where people feel alienated from the decision-making process. Long-term planning requires that the consequences of various courses of action are considered, and that decisions are made based on what is in accord with human values, rather than what is technologically possible.

Perhaps the most important aspect of decision support is that it attempts to link the specialist, who has mastery of a narrow area of knowledge, with the generalist, who has a grasp of the larger situation. In modern democracies, the specialists (often in the public service, which is not directly accountable to the public) take power from the generalists (such as government ministers, who are directly accountable), "because the specialists understand their own particular area of expertise—but not the whole—while the political generalists do not understand the parts and are gradually losing their grip on the whole" (Jones, 1990, p. 175). Computerized decision support systems may be a way to bring specialist knowledge together, so that the whole may be more easily viewed.

GIS is often touted as "an integrating technology", able to link together many data strands into a common structure. GIS, holding the data, must link together users and organizations, by providing support for the decisions that the organization must make collectively. Dynamic modeling, as a method both of simulating the future and of perceiving reality, will be one of the GIS tools involved in this process.

In summary, dynamic modeling in GIS is needed:

- To take GIS beyond being a "data retrieval and sifting engine";
- For GIS to become a decision support tool;
- For providing GIS with other models of reality; and
- To bring generalists and specialists together and remove the alienation between the decision-makers and those affected.

Four-Dimensional GIS

All current GIS can handle 2-D spatial data; this is the minimum expectation of GIS. The limitations of GIS at this level are more the limitations of the particular application's data structure, or the limitations of that particular implementation. For example, a raster GIS provides very little topological structure to assist in network analysis; however, a vector-based GIS may not have a network analysis module, or it may be poorly implemented.

The main limitation in 2-D GIS is that it is trying to model a physical world that is, essentially, 3-D. The effects of this limitation are that elevations are difficult to represent, or are treated as another attribute, rather than as a necessary part of space itself. Being attributes, only one elevation can exist at any one point. Overhangs cannot be represented, and it is not possible to represent things that are buried. For example, two buried objects that are in the same horizontal locations, but at different depths, cannot be represented. This is a major limitation for GIS use in geological
applications, for modeling buildings, and for considering many natural processes that occur at different levels.

**Dynamic Modeling and GIS**

A major problem in linking other applications to GIS is that encountered when trying to link GIS and dynamic modeling. Discussion of some GIS limitations to the integration of GIS and dynamic modeling may be found in Hazelton et al. (1990a) and Hazelton (1991a, 1991b).

Dynamic models that assist environmental management can benefit by being linked to GIS. The GIS provides data collection, storage, manipulation, and presentation facilities for a higher-level management system, of which the dynamic modeling system is another part. A management system that combines dynamic models and GIS would seem to have considerable utility, and it would also allow the large amounts of data associated with the dynamic modeling to be managed more easily. It also allows a GIS to be used as a better tool for decision support.

Investigation of ways of linking dynamic models and GIS has shown that the two systems are not compatible (Hazelton, 1991a, 1991b). There are two areas of incompatibility: 1) Current 2-D GIS cannot handle large quantities of temporally referenced data, such as that generated by dynamic models; and 2) GIS are restricted to working in what is essentially a 2-D environment. Dynamic models for general-purpose use would have to handle 3-D spatial data.

There are two ways of dealing with this problem of incompatibility.

First, one can try to live with it. This has been tried with varying degrees of success. It has required complex linkages and data-exchange methods, and most of the solutions have been ad hoc and not in any way generic solutions. For example, Osborne and Stoogenke (1989) describe a system that allows temporally-referenced data to be used in a GIS, specifically using a model to 'grow' a forest, then using a GIS to analyse the results. That this operation requires two separate GIS (IDRISI and ARC/INFO), the growth modelling system (FIBER, which creates its own temporal database) and a suite of Pascal programs to link them together shows the complexity involved. This approach also requires the use of time purely as an attribute, something that is undesirable when dealing with spatio-temporal data, as it denies the possibility of time having a topological structure.

Secondly, one can alter the components to achieve compatibility. One could try to change the dynamic models to 2-D, but this, naturally, defeats the purpose of the model, and sensibly, no-one has tried it. Far better would be to develop a 4-D GIS and allow it to interact directly with the dynamic modeling system being used.

Two-dimensional GIS have major limitations, if it is wished to incorporate many additional features with GIS, or to extend GIS into new applications. Some of these limitations are being overcome by the development of 3-D GIS (Raper 1989; Pigot 1991).

**Space, Time and Space-time**

Newton-Smith (1986) expressed the view that no satisfactory philosophical understanding of the nature of space, time and space-time can be achieved by remaining at the purely semantic level. Questions about meaning take one to physics, and certain results in physics take one back to meaning. While space and time are apparently well-understood concepts, it is very difficult to explain what they actually are, owing to them being so intimately connected with such a wide range of other fundamental ideas. It is only by considering the use of space and time in a given situation that it is possible to understand what they mean in that context.

Theoretical physics has shown that space and time can be considered to be parts of a single entity: space-time. Most implementations of temporality in databases treat the two as quite separate. Langran (1989b) has reviewed the literature relevant to temporal databases in GIS and discussed the methods available for storing and retrieving temporally referenced data in relational DBMS. These methods considered relatively slow rates of change in the database and had major problems handling a join, or GIS overlay, operation.

As stated in their paper, the methodology proposed by Lan-
gran and Chrisman (1988) relies heavily on a temporal attribute database. All the spatial and topological data in the database's domain are held in one layer, while the topology of the line segments is used to reconstruct the polygonal topology after a time slice is selected. This also may involve a topological 'clean up' for the selection of any time slice. As an alternative, the hybrid GIS modeling system described by Osborne and Stogenke (1989) shows the degree of complexity involved.

Data from different time instants, or epochs, can of course be stored in a conventional GIS, as described in Hunter and Williamson (1990). There are many examples of using GIS to compare the state of things at different epochs. However, a 4-D GIS should be able to store data that represent the complete history of 'real world' objects, rather than just occasional 'snapshots' of the state of things. It also should allow spatial, temporal and spatio-temporal analyses in the same manner as current GIS perform spatial analysis. Dynamic modeling capability is also a necessary component of the entire system. A 4-D GIS can naturally store and process 'non-current' data, i.e., data that represent the history of an object or region.

Hunter (1988) argues the case for retention of non-current data in a manner that allows easy retrieval, and the importance of having these data available. Vrana (1989) discusses the need for recognition of the temporal component of data in LIS, pointing out that the interdependence of the spatial, temporal and thematic attributes of features is not a by-product of such systems, but their very heart. Hunter and Williamson (1990) describe a method of storing data for a digital cadastral database, where the changes in the graphical component were quite slow, compared with potential changes in the attribute database. Here the non-current graphical data are stored in a reference file and 'switched on' when necessary by the attribute database. As the data system was a CADD system linked to a database (Intergraph's IGDS and DMRS, running on an Intergraph VAX), topological structure was not built on the data until a specific instance (time-based) was selected.

Hazelton et al. (1990b) proposed a classification of temporal GIS, based on the expected rates of change over time of the graphical and attribute components of the system. Where the attribute database changed more frequently than the graphical database, the GIS was termed 'Slow', while the reverse situation was termed 'Fast'. A digital cadastral database would therefore be 'Slow', while an environmental monitoring GIS may well be 'Fast'.

Hazelton et al. (1990a) discussed a simplified version of a spatio-temporal GIS data structure. The essence of this data structure was to treat the history of 'real world' objects as 4-D objects, and to develop a 4-D GIS to handle these types of objects. Many aspects of this conceptual GIS data structure are developed by analogy from 2-D GIS. This GIS would be a vector-based system, rather than a cell-based system.

Bell et al. (1990) describe the basic approaches they intend to follow to create a 4-D cell-based GIS. The aims are similar to Hazelton et al. (1990a), but the data structures are simpler. Both these approaches have a place in the range of 4-D GIS, as cell-based and vector-based systems have their roles in 2-D GIS.

Langran (1989a, 1989b) and Langran and Chrisman (1988) can be said to sum up the situation of spatio-temporal data structures as follows:

- Spatio-temporal indexing of data is a major problem;
- Structures for storing spatio-temporal data will not be simple;
- There is unlikely to be a single spatio-temporal data structure that will be widely acceptable or useful.

However, despite some use of GIS to store temporally referenced data, there does not seem to have been an implementation of a true 4-D GIS to date.

Integrating Dynamic Modeling and GIS

In the preceding discussion, we have argued that the most general solution to integrating dynamic modeling and GIS is the development of a 4-D GIS. This leads to questions about the role of dynamic modeling in a 4-D GIS, and how the two systems should work together.

If we consider using a straightforward cell-based GIS such as GRASS (USA-CERL, 1988), and we wish to use a purely Boolean empirical model, we can use the module 'com-
bine'. If we wish to use a 'graduated' Boolean empirical model, we can use the modula 'weight'. If we wish to use a regression model, we can use the module 'Gmapcalc'. Such operations can be done in various other ways in other GIS, but the facilities generally exist as part of the basic GIS package. The input required to run the model is 2-D data, a model and its parameters; and the output is further 2-D data.

If we use a dynamic model to generate future scenarios of a system, the result is a series of 'time-slices', each of which shows the state of the system at a particular instant of time. Each time-slice may include several attribute 'layers' or 'geographical extents'. The input required for this type of model is several 'time-slices' of 2-D data (to allow calibration and verification), a model, its rules and parameters; the output is several 'time-slices' of 2-D data. This provides a simple means of telling the difference between the two types of modeling in a conventional, 2-D GIS.

However, in a 4-D GIS, the input and output of any type of modeling is rather different. An empirical model in general will take some 4-D data and a model, and produce 4-D data. A dynamic model will also take 4-D data and a model, and produce 4-D data. The dividing line starts to blur.

Calkins (1988) produced the following taxonomy of GIS uses, which contains six groups. These are, from highest to lowest usage:

1. Inventory
2. Map-based Presentations
3. Query and Map-based Presentations
4. Simple Map Comparison
5. Complex Spatial Modeling
6. Spatial Decision Support

At the lowest level, Inventory, the GIS requires no capabilities other than storage and retrieval. At higher levels, more modeling in general, and more complex modeling in particular, is introduced. Level 3 requires the use of empirical models at a simple level, and probably indicates the lowest level of what could be clearly stated to be a 'proper' GIS, rather than a digital mapping system. Level 5 requires the use of dynamic models, and as such is going beyond the capabilities of most GIS on the market. Level 6 goes beyond what we currently refer to as 'GIS', as in a Spatial Decision Support System we could expect GIS to be just another tool available within a much broader application system.

If we consider the use of GIS at Levels 3 and 4 (where most 'true' GIS operations are performed) we notice that empirical modeling is such a necessary part of GIS that it has been subsumed into the GIS. Average users think about using the GIS to find the most suitable housing sites, for example, rather than thinking about using a model to find them. However, the same users think about a dynamic model as being quite separate from the GIS, as this is their usual experience. That this discussion has presumed various levels of GIS usage are 'true' or 'proper' GIS operations, is an indication of this conceptual separation.

In a 4-D GIS, we may expect that dynamic modeling will become part of the essential fabric of the GIS, as empirical modeling has become in 2-D GIS. The definitions of modeling in 4-D will have to be revised, both because of this incorporation and the differences between empirical models and models based upon 'fundamental laws.'

**Conclusion**

In this paper we have argued that GIS is severely constrained by its dependence upon the cartographic metaphor as its model of reality. Development of GIS beyond a 'map representation' system requires GIS to become a 'reality modeling' system. At the most basic level, GIS must be able to handle truly 3-D spatial data, and temporal data.

With the ability of GIS to incorporate time on the same basis as space, dynamic models that use both time and space as parameters can be used to model the consequences of various management options in many application areas, such as agricultural and environmental management. Dynamic models in general will require the management of spatial data that is truly 3-D, over time. This will necessitate the use of a 4-D GIS.

The extension of GIS beyond the current 2-D design will require re-thinking of the relationship between modeling and GIS. In the same way that 2-D GIS has subsumed empirical
modeling, 4-D GIS will subsume dynamic modeling.

At its most fundamental, GIS has been considered to be a decision-support system that handles spatial data. At present, the decision-support role is limited by the inability to produce and analyse future scenarios in any depth or quantity. Four-dimensional GIS provides the basis for development of better systems for decision support.

Acknowledgements.
This work has been primarily sponsored by Intergraph Corporation, under the supervision of Prof. Ian Williamson. Andrew Turk, Walter Hesse and other researchers in the Department of Surveying and Land Information have been involved in many discussions with the authors about these issues, and have contributed to the development of the ideas expressed in this paper. Marcus Lester, Simon Pigot, John Herrin, Tom Watson, Gail Langran and Nick Chrisman also discussed a number of the issues involved in this paper and the background to it.

Note
1. This is akin to Gell-Mann's totalitarian principle in quantum mechanics: "whatever is not forbidden is compulsory," Fitzgerald (1970).

References
Bell, S.B.M., Chadwick, A., Cooper, A.P.R., Mason, D.C., O'Connell, M., and Young, P.A.V. 1990. "Handling Four Dimensional Geo-Coded Data." Proceedings of the 4th International Symposium on Spatial Data Handling, held in Zürich, Switzerland, August, 1990, pp. 918–927.
Calkins, H.J. 1989 Seminar on GIS, Department of Surveying and Land Information, The University of Melbourne, Australia.
Falls Church, Va.: ASPRS, ACMS.
Systematic Development of Education and Training Programs: A Key to Successful GIS Development

Charles R. Blinn, LLoyd P. Queen, Lorentz R. Hegstad and Daniel J. Fitzpatrick

Charles Blinn is an associate professor and extension specialist in the Department of Forest Resources at the University of Minnesota.

Lloyd Queen is a research associate in the Department of Forest Resources at the University of Minnesota.

Lorentz Hegstad is assistant to the director, Minnesota Department of Natural Resources, Division of Forestry in St. Paul, Minnesota.

Daniel Fitzpatrick is a graduate research assistant in the Department of Forest Resources at the University of Minnesota.

Abstract: As GIS technology is adopted by organizations, consideration and effort should be directed to strategic planning and design issues regarding system development. Often these efforts take the form of systems analysis or user requirements analyses. This paper establishes a framework for similarly assessing the educational and training needs of staff and project personnel. Education and training are fundamental to the successful implementation of any GIS project. The process of assessing educational needs is termed a learner-needs assessment. A case study of a learner needs assessment for a large state agency in Minnesota is presented to demonstrate the application and utility of such an approach.

The geographic information system (GIS) and land information system (LIS) development process includes planning, analysis, design, evaluation, and implementation. (For the purposes of this discussion, we will use “GIS” as an inclusive term containing “LIS” and equal to “GIS/LIS.”) Throughout this process, the focus is traditionally on technical issues related to hardware, software, and database concerns. Organizations typically perform a User Requirements Analysis (URA) to define guidelines, specifications, and standards for GIS technology implementation (Antenucci et al. 1991; Antenucci and Levinsohn 1990; Gupptill 1988 and 1989; Montgomery 1989; and Somers 1990). The URA planning strategy will often consist of several chronological phases, including user identification and interviewing, data and map inventories, and database and system design. As all organizations are unique, so too are their GIS needs.

Although the goal of the development process is to design a system that satisfies the organization’s needs, it typically focuses on technical issues only. Though an organization may have the best available system design, implementation may be impeded if users are not prepared to adopt the technology. The human resource needs of system-users and end-users require the same systematic consideration given to the technology. It is the authors’ thesis that staff development, accomplished through education and training, will help identify and meet those human resource needs.

GIS education addresses basic concepts such as: definition of a GIS, how to manage data with GIS, and example applications. It also addresses management...
issues including: organizational commitment, administration and staffing, implementation timelines, and organizational policy. Since education includes a focus on attitudes, motivations, and behavior, early education can help shape the feasibility studies, the URA, and system implementation. Education will also increase the user's awareness, stimulate interest, and enhance their ability to evaluate the technology. In contrast to education, GIS training typically provides hands-on learning experiences with hardware and software, generally focused on a specific system. Although organizations frequently provide numerous training opportunities to selected staff, they often overlook the educational needs of the organization as a whole. While there are several possible means by which organizations can obtain education and training (Table 1), only a program that is systematically designed and implemented by that organization can address the diversity of audience needs unique to that entity (Onsrud and Pinto 1991; Sullivan and Miller 1991). Education can thus be incorporated into the development process at an early stage while training will occur once a system begins to be implemented. The goal of our paper is to overview the systematic education planning process with particular emphasis on assessing the needs of targeted audiences. A case study that focuses on conducting a learner needs assessment is presented to elucidate these concepts.

**A Systematic Methodology for Planning Educational Programs**

Although high-level decision-makers will provide oversight to the development of an education and training program, a subcommittee of organizational members can provide substantial guidance on a continuing
basis. That subcommittee should contain members from each department or division representing the distinct audiences that will be impacted by the GIS. Membership on the committee of a professional educator is recommended. The steps involved in the process of developing a GIS education and training program include:

1) Establish program goals.
2) Identify target audiences.
3) Assess the needs of each audience.
4) Establish specific learning objectives for each audience.
5) Design learning experiences for each audience. Learning experiences refer to the educational activities which will result in changes in knowledge, skills, or attitudes on the part of the audience (Boyle 1981; Tyler 1949).
6) Conduct evaluations on an ongoing basis.

Identify Target Audiences

An education and training program should target audiences to ensure that the unique needs of diverse audiences are addressed by relevant learning experiences. While this may seem obvious, too frequently programs attempt to address all audiences simultaneously with the same learning experiences.

The purpose of targeting audiences is to identify those groups or classes of individuals who share similar backgrounds and information needs and who therefore share similar educational needs. Within a targeted audience, members will frequently have similar educations, work at similar levels within the organization, and have similar professional goals or aspirations. It is also the case that most members of a target audience will have advanced to a similar stage of innovation adoption. A model of innovation adoption (Table 2) was proposed by Rogers (1962) and later refined by Rogers (1983) and White (1975). Rogers’ model of the stages of innovation adoption can be a useful guide in recognizing important characteristics of specific target audiences based on their level of innovation adoption.

Once the target audiences are defined. Steps 3, 4, and 5 in the program development process should proceed separately for each targeted audience. Preferably, one or more members of the target audience should participate in each remaining step of the planning process.

Assess the Needs of Each Audience

In designing learning experiences, educators need an understanding of the interests and backgrounds of the audience so that appropriate experiences that will evoke learning can be designed (Boyle 1981; Knowles 1980; Tyler 1949). Furthermore, the effectiveness of the learning

---

**TABLE 2.**

Stages of innovation adoption, as proposed by Rogers (1962).

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Awareness</td>
<td>Mass communication channels promote awareness. The potential adopter is exposed to the innovation, but is not yet motivated to seek additional information.</td>
</tr>
<tr>
<td>2. Interest</td>
<td>The potential adopter favors innovation in a general way, but has not yet judged utility in terms of their own situation. They seek additional information; norms of social system affect where information is sought. Negative attitudes can be formed at this stage.</td>
</tr>
<tr>
<td>3. Evaluation</td>
<td>Information from peers is sought. The potential adopter mentally applies innovation to present or anticipated future situation. “Mass communication” is too general to provide reinforcement.</td>
</tr>
<tr>
<td>4. Trial</td>
<td>The potential adopter uses the innovation on a small scale to determine utility in their own situation; personal experimentation. They seek information specific to the methodology of the innovation. Rejection of the innovation can occur if results of personal trial/experimentation are misinterpreted. This stage may be skipped by late adopters.</td>
</tr>
<tr>
<td>5. Adoption</td>
<td>The potential adopter decides to continue with full use of the innovation.</td>
</tr>
</tbody>
</table>
experience is directly related to the degree of involvement of the learner in its determination, implementation, and evaluation.

Audience needs can be determined through a "learner needs assessment," which refers to a systematic process for collecting and analyzing information about the educational needs of individuals, groups, or organizations (Moore 1980). This assessment is ideally conducted by the educator(s) who will be providing the education or training to ensure that information resulting from the assessment is familiar to them as they design and deliver the learning experiences.

During the needs assessment, the target audience should be questioned to determine:

1) What they already know and what they need to learn about GIS.
2) What attitudinal issues should be addressed.
3) What motivational factors are most important (e.g., under what conditions do they learn best, preferences for how seminars or formal workshops might be structured, who does that delivery).
4) Which other individuals should be involved with the development and delivery of the program.
5) Long-term and parallel needs for education and training.

Answers to the above questions will help in identifying the stage of innovation adoption (Table 2) that best describes the audience's current relationship to the goal of GIS adoption. The answers will also enable the educators to set goals for moving the audience into higher-order stages of adoption.

The needs assessment should identify which types of learning experiences are likely to be the most effective for each audience. The learning experiences may include a blend of formal seminars or workshops with informal tools such as videos, newsletters, tutorials, electronic bulletin boards, and user groups. Informal learning tools are an important means of reinforcing information presented in the formal settings. The stage of an audience's adoption of GIS will influence how best to address needs (Rogers 1962, 1983).

An assessment of learner needs can be accomplished through a variety of methods (Robinson 1979). A questionnaire/survey, focus group interviews, or a combination of these methods should be appropriate in most cases. The summarized findings from the learner needs assessment represent the basis for establishing specific learning objectives for each audience.

Establish Specific Learning Objectives for Each Audience

After determining the needs of each targeted audience, specific learning objectives for each audience should be established. These objectives should address the overall program goals and will be reflected in the design of the learning experiences (Robinson 1979). They become the criteria by which educational materials are selected or developed, content is outlined, and evaluations are conducted (Tyler 1949). The integration of the individual audience objectives into the broader program goal of GIS adoption will be accomplished by moving each audience from their current stage of GIS adoption to the next stage (Table 2).

Two sets of outcomes should be established for each audience to guide the program evaluation (Step 5). First, establish the overall broad goal that will be fully attained after conducting a series of learning experiences. This goal is necessary to provide a clear, long-range direction for the effort. Second, identify short-term objectives that will help guide development of learning experiences during the intermediate process.

Design Learning Experiences for Each Audience

When designing learning experiences, consider the primary focus (individual, departmental, or organizational) of the experience, setting or format in which learning will take place, instructional resources (length, speakers, instructional materials) and aids (video, slides, workbooks), placement of review sections, sequence of units, and number of learners. Seminars and workshops for complex subjects such as GIS benefit from the use of several complementary approaches. These approaches might include: presentation of theory through readings and lectures, demonstrations, practice by the learner, and continued coaching by the instructor (Houle 1972). This strategy encourages participation, helps to maintain the interest of learners.
and instructors, and considers the differential response of individuals to various approaches. The model for the stages of innovation adoption (Table 2) can provide evidence on which to base the selection of complementary approaches.

Conduct Evaluations on an Ongoing Basis

Evaluations should be conducted to determine the effectiveness of learning experiences, to ascertain program status, and to identify areas for instructional improvement. While “exit poll” evaluations at the end of a formal learning experience can provide immediate feedback, they seldom assess the degree to which changes in knowledge, skills, and attitudes have actually taken place.

Evaluations should be designed to measure the degree to which identified program goals (Step 1) and objectives specific to the individual learning experiences (Step 4) have been achieved. Evaluation of an overall program should appraise knowledge, skills, and attitudes as reflected in the movement from an earlier stage to a later stage in the adoption of GIS (Table 2). An assessment should be made early in a program, with follow-up appraisals conducted weeks or months after learning experiences are provided. Evaluations conducted several months after completion of a learning experience will help provide evidence of longer-term changes in knowledge, skills, and attitudes while also serving as the basis for subsequent learner needs assessments and future program planning.

The evaluation process should be planned at the outset of the program to ensure that program status can be accurately assessed throughout its life cycle. Before conducting any evaluation, a written plan that defines what is being evaluated, evaluation audiences, the purpose of the evaluation, methods used to collect and analyze data and evaluation tools (questionnaires, interviews, focus groups, etc.) is necessary. Successful evaluation requires the support of high-level decision-makers and all other organizational members who will be affected by it. Information about survey design may be found in Dillman (1978), Patton (1987 and 1990), Rossi et al. (1979), and Stecher and Davis (1987). Standard evaluation protocols have also been established (Joint Committee on Standards for Educational Evaluation 1981).

Case Study

Use of the described methodology will facilitate the development of successful and effective GIS education and training programs. These methods were employed during the development of a GIS education program for the Minnesota Department of Natural Resources (DNR), a large state agency in Minnesota. This structured education program planning process was part of a broad effort on the part of the organization to adopt and advance the use of GIS technology. The content, organization, and conduct of that program, with emphasis on the learner needs assessment, is the focus of the next section.

Background and Organizational Context

The DNR’s mission is “To preserve, protect and enhance Minnesota’s natural resource heritage in order to benefit the environment, economy, and quality of life for all Minnesotans, present and future” (Minnesota Department of Natural Resources 1991). The central administration is located in St. Paul and each of the six regions is administered locally by a regional administrator. Because of the distribution of natural resources across the state, the 14 DNR divisions and service bureaus are not equally represented in each region. There are about 2,500 authorized full-time equivalent (FTE) employees in the DNR.

The DNR completed a geographic resource information system (GRIS) study in 1991 which followed a URA approach. Because the DNR has broad interdisciplinary responsibilities, GIS technology is being adopted at many levels within the organization. In addition to obvious program applications (i.e., inventories of forest roads, public access to lakes, recreation trails), it is anticipated that GIS will have broad impact on many processes within the DNR, including departmental planning, environmental impact assessment, and integrated resource management.

There are currently about 15 system-users with substantial GIS experience within the DNR. There are a larger number of
end-users who have been involved with the development of individual projects. While some system-users have received extensive education and training, others were self-taught. Many of the end-users who will be involved with GIS have received some ad hoc or meeting-associated GIS education or training.

In 1989, the Minnesota Extension Service funded the design of an educational seminar that was to be focused on the DNR; the DNR agreed to fund the delivery of that program and to commit staff time to that effort. The seminar was to be part of a broader program that the DNR would be responsible for developing and implementing. As the DNR had not yet fully conceived that program, some of the program-planning steps were not completed in great detail. However, both organizations were committed to the design and delivery of the initial educational seminar.

**Approach**

A design team, comprised of the DNR’s GIS/LIS Working Committee and the first and second authors of this paper, was formed to initiate the program planning process. Those authors had been involved with several previous computer-related education and training programs for DNR staff as well as research projects that involved DNR personnel. The GIS/LIS Working Committee is comprised of central administrators from several divisions and bureaus.

The following objective was established for an overall GIS education and training program:

To facilitate the adoption of GIS technologies within the DNR, wherever those technologies can help the DNR to serve the public better by managing Minnesota’s resources more efficiently.

The design team decided to focus their efforts on one target audience, non-supervisory resource management professionals, because financial resources were not available to develop a full-scale education and training program. For the most part, those professionals include resource managers who will use GIS products to assist on-the-ground decision-making (e.g., district foresters, park managers, area wildlife managers, mine-land reclamation specialists). That audience was selected because they will become the end-users, they are not centrally located, and they had not all received basic education about GIS. Within the DNR, there are about 750 individuals in this target audience. DNR central administrators were excluded as they had already been the target audience of several GIS education programs delivered by other cooperators.

A comprehensive set of questions that comprised the scope of the assessment were developed by the design team. Questions were then organized into an outline that was used to initiate group discussions (Table 3). It was decided that the best way to conduct the assessment was through a series of focus group interviews (Krueger 1988) conducted at the regional level.

The learner needs assessments were designed to last three hours. Three regional administrators were requested to organize a session that contained members of the target audience from as many divisions and bureaus as possible. It was suggested that including one or two staff members with GIS experience might be useful. Focus group size was limited to 10 people. The agenda for the meeting, conducted by the first and second authors of this paper, included an overview of the purpose for the meeting, a short education session that presented some basic GIS concepts, an overview of the framework for the survey, and the needs assessment itself. The actual needs assessment survey occupied two-thirds of the available time.

In addition to the focus group discussions, each participant completed a written questionnaire that sought their input on course content and delivery options. This questionnaire was developed to be as comprehensive as possible, based on the design team’s knowledge of the target audience. It contained the following items:

1) Nine potential main sections that might be included in an education program (Table 4). Each main section was posed as a question and included several sub-questions. Each sub-question represented a potential topic, area that could be addressed in the learning experience (e.g., for the main section entitled “What is a GIS?”, there was a sub-question “What kinds of data are in a GIS?”). As the questionnaire was not comprehensive, space was provided for participants to pose other questions either as a new main section or a sub-question. Participants were asked to rate each sub-question using a Likert scale.
TABLE 3.
Main survey questions included in the education learner needs assessment of the Minnesota DNR.

A. Reactions to new technologies
   1. What have been your experiences with new technology?
   2. What experiences did you have when you were first introduced to computers?
   3. How would you rate your knowledge about GIS?
   4. What do you know about GIS implementation in the DNR?
   5. What are your concerns about implementing GIS?

B. Development of the education program
   1. Describe a good educational program.
   2. What would stimulate you to attend a voluntary program?
   3. What do you hope to get out of this course?
   4. Evaluate a draft educational program and delivery options.
   5. What are your concerns about being part of an audience comprised of different Divisions and Bureaus?
   6. Who else should be involved with the presentation of this GIS educational program?

C. How to develop a broader educational initiative
   1. What are other types of educational information about GIS that should be made available to you?
   2. What should be the format of any long-term commitment toward GIS education and/or training?

D. Additional thoughts, concerns, or feelings that would help us better understand your needs

(Babbie 1982) with a range of 1 (least important) to 7 (most important).

2) A ranking of the main sections included in Item 1 (described immediately above). This allowed the evaluators to determine which sections were most important so that particular emphasis could be placed during the education program. This also allowed the consistency of responses (between priorities indicated in Table 3) to be evaluated.

3) A ranking of seven options for delivering the program (at annual training meetings for specific disciplines, by video, satellite downlink, discipline-specific—once per region, interdisciplinary—once per region, discipline-specific—many times per region, or interdisciplinary—many times per region). Additional space was provided for participants to enter and rank other delivery options.

DNR central administrators from the design team were also asked to complete the written questionnaire. They were asked to consider the needs of the target audience in their response.

Results

Twenty-six people, representing eight divisions and bureaus, were interviewed during the learner needs assessments conducted in the three regions. Additionally, five central administrators returned their completed surveys. In general, participants were pleased to be part of the initial design of the education program. Many indicated a belief that the needs assessment was an appropriate first step in the design of the program.

Learner needs assessments responses were summarized and submitted to the design team. As a result of the needs assessment, many important points were learned that might have been overlooked or not adequately emphasized if the assessment had not been conducted:

1) There was at least as much concern for GIS management and policy issues as there was for basic technical concepts.

2) Key high-level DNR staff members should present part of the program, including addressing many of the management issues.

3) Support networks within the organization should be established. These networks could include new GIS staffing, telephone support, newsletters, and local user groups.

4) While GIS has the potential to support interdisciplinary decision-making and natural resource management, the focus of the program should be to get the individual user involved. Once individuals become involved, interdisciplinary approaches to problem-solving will evolve.

5) Learning experiences that make use of multiple media and forums should be developed. Offering learning experiences on a continuing basis for audiences at different stages of adopting GIS (Table 2) were strongly encouraged.

After careful review and discussion of the summary report, an educational seminar was designed and developed. Based upon the learner needs assessment and the recommendations of the design team, it was decided that the seminars should be presented to multidiscipli-
nary audiences composed of several divisions/bureaus. The specific seminar objectives were: To demystify GIS, to let you (DNR employees) know where you “plug-in” to GIS, and to present the DNR’s vision for GIS.

Time and travel constraints suggested that the seminar should be repeated at multiple locations. Nine one-day education seminars were conducted between December 11, 1990 and February 7, 1991. The seminars were presented by two university staff members and three DNR personnel. University staff helped to define GIS and to provide an outside perspective on how the technology can assist in solving natural resource management problems. DNR presenters helped to validate the application of the technology and to indicate the DNR central administration’s support for GIS and for the development of staff to become involved with the technology.

A written “exit poll” evaluation was conducted at the end of each seminar. That evaluation indicated that the seminars met audience needs. This was reinforced through informal discussions with several audience members after each session. Many attendees indicated that they were pleased that their needs had been considered in the design of the learning experience.

An evaluation to measure the attainment of stated program objectives will be conducted in 1992. An additional purpose of evaluation will be to gain insight into the status of the diffusion and adoption of GIS technologies within the DNR. This information will assist in the continued development of an integrated and comprehensive educational program to facilitate the adoption of GIS within the DNR.

Summary

Before implementing a GIS or any other information technology, organizations should conduct targeted education programs that address both basic GIS concepts and management issues. Training programs should be conducted when the system is delivered for personnel that will actually become system-users.

Every organization has its own unique set of staff needs. While a user requirements analysis is conducted to determine specific system specifications for current and future needs, an organization should also follow a structured planning approach to develop an education and training program for its staff. That approach should establish overall program objectives, identify target audiences, assess the needs of each audience, establish specific objectives for each audience, design learning experiences for each audience, and conduct evaluations on an ongoing basis. The end result should be a more successful GIS/information technology implementation.

The Minnesota DNR is now prepared to move forward with system design and analysis efforts that are newly coupled with a better awareness for staff development needs. This structure, made possible through the tandem use of a URA and this systematic process for evaluating education needs, will help to make adoption of GIS technology more successful throughout the organization.

Notes
1. System-users are defined here as those who use GIS “hands-on” while end-users are defined as those who use information output from GISs.
2. For the purposes of this discussion, we refer to education as “the process of successful learning of knowledge, skills and attitudes . . . usually (in contrast to training) where it is learned in such a way that the learner can express his/her own individuality through what he/she learns and subsequently apply it, and adapt it flexibly, to situations and problems other than those he/she considered in learning it” (Rowntree 1982). Training refers to “the systematic development in a person of the knowledge, skills and attitudes necessary for him/her to be able to

---

TABLE 4.

Potential main sections of an education program that were evaluated by participants in the Minnesota DNR learner needs assessment.

| A. What is a GIS?  |
| B. How does GIS work?  |
| C. What can a GIS do? Not do?  |
| D. How do I actually use a GIS?  |
| E. How is the DNR planning to use GIS?  |
| F. Who else is using GIS? What are they using it for?  |
| G. What can/should I do to prepare myself for using GIS?  |
| H. What is the benefit/cost of using GIS?  |
| I. What is the DNR’s long-term commitment to GIS?  |
perform adequately in a job or task whose demands can be reasonably well identified in advance and that requires a fairly standardized performance from whoever attempts it" (Rowntree 1982).

References


EDITORIAL INTENT

The *Urisa Journal* was originally conceived as having several sections. The Refereed portion would be the most intellectually rigorous, suitable for academic submittals. The Feature section would be more journalistic and visual (that is, giving special emphasis to maps, photographs, other graphics). The remaining sections—Feature Map, Reviews, In My Opinion—are, hopefully, unambiguous. Now, with the Journal beginning its fourth year, this conceptual framework appears to be working. We will, accordingly, carry on.

In the Features section, we welcome material that is: serious or not so serious; visionary or pragmatic; domestic or international; public or private sector; high or low tech; far-fetched or down-to-earth; managerial or technical; vocational or educational; qualitative or quantitative; GIS or non-GIS. In short, we are open to different ways the world can be explored, described, understood.

Norman Cousins has said "No one really knows enough to be a pessimist." Assuming further that no one can ever know enough to be a pessimist, the pursuit of knowledge must therefore be an optimistic (or, at worst, realistic) endeavor. Torturing the logic a bit further, an optimistic endeavor is joyful and therefore—ah ha!—contribute material to the Features section is pleasurable. So, please do so.

We only ask that you submit material that you genuinely believe will interest URISA's members, and of which you are truly proud.

*Gilbert H. Castle, III*
*Warren Ferguson*

In this issue... we begin with transcripts of two presentations delivered at the annual conference of the Australasian Urban and Regional Information Systems Association (AURISA).

John Keene, in his humorous but biting style, argues that bureaucratic ineptitude is crushing the GIS industry, "Down Under." His observations hold equally true for parts of North America.

Will Craig, a past President of URISA, illustrates recurring problems in assembling digital data bases. "Illustrates" is used here literally; cartoons and graphics are used to punctuate his points.

Laurel McKay, URISA's President-elect, describes the key events leading to the creation of Land Information Alberta—one of the world's largest, most innovative, and publicly accessible urban and regional information systems.

Finally, in what turns out to be a Journal issue with definite presidential overtones in this election year, the time-honored "Short Rules" of URISA's founder and first president, dear Horwood—are resurrected for us by Steve Kinzy.
GIS-ing About in Public: An Anatomy of a Public Disgrace
John Keene

Over recent years, much has been written and spoken at such talkfests as AURISA about Public Enquiry Systems for Government information. The basic concept, which is part of official government direction in a number of Australian states, is that the data held by government are made available on-line to the public for a fee. The service is aimed at conveyancers, surveyors, valuers planners, and a variety of developers, greenies, and other interest groups. The fee is intended to cover the service itself, the data integration and further data capture—representing a modest return on investment. Many states, in this era of privatization, believe that this service is best provided by the private sector.

Genasys, as a company, has been interested in such schemes for over three years. It has spent about $2 million on market research, marketing and prototyping. Three years ago, Australia was an industry leader in this area and had the opportunity to develop a major export industry. This opportunity has now virtually vanished. What happened? Why? What went wrong? Can we right it?

Let us analyse what has led to this situation:

1. **Information Control**
The basic data are held and owned by the government. They were collected for a variety of purposes over the past two centuries. While government has the control of the data, it does not have the data under control. The information has little data integrity. Up to 20 percent of line items in a given data set can be wrong—as found when different data sets are matched against each other. There is considerable work (and cost) in correcting the data. Individual departments feud over who “owns” the data and who gets the revenues. (The answer of course is that the government owns the data, not any given department.) The most appropriate metaphor to describe this conflict between departments over data is two guerilla forces sniping at each other over a piece of waste swamp.

2. **Policy Position**
In the policy, there are no precedents. Government is uneasy where there are no rules to follow. Creativity, courage and original thinking are as prevalent in government as they are in sheep paddocks. It takes focus and commitment and the ability to tread where none has trod before. When the crunch of a decision comes, the bureaucrats (with their penchant for reglamania) head for the hills. There are some genuine reasons for this. Public good cannot be measured in dollars and cents—and government holds a clear mandate and responsibility for information dissemination with a public good component. Clarification of policy issues to enable government to move beyond their current functions requires hard work—not simply a commitment to the private sector and some good will.

3. **Lack of Market Perception**
Governments seem not to understand the commercial marketplace. Market research, which identifies timeframes, value-added services required, and government financial returns, are simply ignored. The requirements for information and delivery for the public are met with the arrogance of power—there is virtually no sensitivity to what the public (i.e., voters who use information) require. A three-year decision process is difficult to justify within normal commercial timeframes. There seems to be no understanding that the three key strategic projects in which Genasys is currently involved affect the lifeblood of an Australian industry and are not isolated commercial deals.
4. Government Limitations
There are some things that government does adequately to well. They do not include marketing services, support of clients, funding enterprises flexibly and responsively, or the provision of skilled personnel (technically) to get an operation off the ground, on-time, within budget. The perception is probably correct that this is better handled by private enterpris—or would be—if only the government could work out a model or policy for its relationship with private enterprise. Over a three-year period, no state government in Australia has been able to do this for land information purposes—and yet it has been done in Canada, the United States and Europe during this period.

5. Structural Instability
Change is obviously an occupational hazard during any protracted period of contract negotiation. However, it provides a built-in structural instability. In NSW alone, for example, since we began negotiating with the government, we have changed ministers and minders, departments, head of information services and sundry luminaries. We have had 18 people, including six external consultants, sit on three separate committees with 12 government employees representing seven government departments. All have signed off the proposal but the submission is still being reviewed. Of these people, two have retired hurt (pregnant), one has taken voluntary redundancy, several have changed jobs and one has died. Our total estimated cost to government is over $1 million, and still no decision. Our business partner (IBM), which is really a private enterprise government department, has had over 20 people as “key” to the decision process. Each other state government, including Victoria and Queensland, has a similar litany of structural instability that prevents the job from being done. Someone has likened this situation to swimming through lakes of peanut butter.

Over recent years, accusations have been aired over political and financial scams in government. Various governments now are cleaning up their acts. Any sniff of inappropriate behaviour, and heads will roll. This, of course, is proper and appropriate. However the consequence is that in modern state government the decision process is so much reviewed that it results in proposals being frozen into immobility. Due process takes years. This delay for review has always existed but much more so in recent years. What we have is the triumph of procedure over content.

Conclusion
So what can we do? Perhaps government is a disease with no known cure. Clearly, private enterprise must understand the governments’ cause for concern. The process will take time, and the private enterprise will need deep pockets to keep teams together for projects while the government resolves these issues.

Government, in its turn, must understand that land information utilities are high-investment, long-term projects—returns are only possible in five to 10 years, and only with a project partnership ethos from government. Despite the situation, projects can be funded by the private sector but they need cooperation and commitment from government to get the project done.

In Greek mythology the gates of Hell had written over them “Abandon hope all ye who enter here.” I hope that it will not be the title of next year’s paper.

John Keene is director of Geanys, a GIS software company.
Why We Couldn’t Get the Data We Wanted
William J. Craig

In November 1991 I was a keynote speaker at the AURISA conference in Wellington, New Zealand. My talk was based on an article I co-authored with Paul Tessar and Niaz Ali Khan in the Spring 1991 URISA Journal, “Sharing Graphic Data Files in Open System Environment.” Sharing is an important issue for two reasons: to reduce the cost of data capture and to gain the benefits of synergy, the result from combining data sets. We found that the biggest impediment to sharing data was the relatively low quality of the original data files; people cut corners to meet their own needs, whereas a little additional effort would have created files that met everyone’s needs.

The message was clear, but the presentation was too dry. I contacted a local commercial artist (friend), Scott Fraser, to help me add some spice. Together we made a slide show. The drawings that followed flowed from Scott’s pen and his Macintosh, using Aldus Freehand, 3.0. The issues we addressed, and the text, originated from the Minnesota Data Exchange Project and from the literature review for the original article.

By putting the problems in graphic form, it became easier for the uninitiated to see and understand the issues. Veterans enjoyed the fun, even as they grimaced and identified all too well with the problems.

From listening to people, it’s clear that the problems we encountered are only the tip of the iceberg:

- An Australian told me of getting files with full coordinates, eastings and northings, but no clue about which UTM zone they were in.
- A local county official told of getting a 200-foot-long straight feature digitized by a photo interpretation company, with a vertex every two feet: “It’s easier to use stream mode digitizing, because then the staff doesn’t have to think.”
- I’ve heard about digitized street maps where the operator left gaps, in order to have a place to insert the street names.

The text that accompanies the graphic is based on the script from that slide show. I am willing to make copies of the slide set for anyone who wants one. The originals are in color, colored lines on a black background for Freehand graphics. The cost of reproduction, sales tax, shipping and handling is $23 for the set of 29 slides.

Make checks payable to the University of Minnesota. My address is Center for Urban and Regional Affairs, 330 Humphrey Center, University of Minnesota, Minneapolis MN 55455.

Unusable material or format problems
Our man is being handed a tape, which he cannot read on his microcomputer. Langford (1987), a researcher attempting to build a comprehensive environmental database in Alberta borrowing from existing computerized files, talked about receiving tapes in the wrong density and even paper printouts.
I wish *it were* Greek to me. Ever get a file that was total gibberish? We did. Not even the characters were recognizable, let alone the paragraphs.

I have no idea what the kanji characters mean. If the words had been spelled out in Roman characters, I would know how to pronounce them. Then I could phone a friend and ask for a translation.

*Computers can't read* graphics. We know of one case (Bretano 1988) where the recipient was sent a *symbolized* file. He expected a continuous roadway, but he got a set of unconnected dashed lines. Those dashes are useless in a GIS and he had to request a new file.

**Uncommon Classification**

Sometimes the organization creating the data file had a different perspective on what was important and classified objects differently from what we expected.

Idiosyncrasy and idiot are words sharing a common Greek word meaning ‘peculiar.’ Maybe it’s time to look for a little more commonality.

**Mapping Imperialism.** Here’s a version of the classification problem where a member of the wood products industry has called all lands having 10 percent or more crown cover, forested. Most of us would need a 50 percent cover before we would be convinced that we were in a forest.

**Cost Problems**

One translator we wanted to try, ARC/SIF, cost $10,000; this is a reasonable price for something which would be used on a regular basis, but ridiculous for a one-time effort. Of course there was no offer of any short-term lease.

Translation is always more frustrating and time consuming than one might expect, and staff time costs money (Hultquist and Scripter 1987). Trying to transfer data by modems can run up huge long-distance telephone bills (Langford 1987).
Charging for data, a fact that seems to have been carried to excess in some parts of Australasia, can severely reduce data sharing. I talked to officials from a state department of transportation, from a state natural resource department, and from a city government who could not afford to acquire the state-created parcel map to use as a base for their mapping. Their two options are 1) to drop GIS altogether or 2) to build their own base map. In either case, the state loses the benefits of establishing a multi-purpose synergistic database. Aren't they all supposed to be working for the same taxpayer? Who's paying the bills, anyway?

Translation Problems
Not all translators are created equal. These guys all look like they could do the job, but looks can be deceiving. Some are older or in poor health. Some are not so good at numbers. Sometimes the one you need isn't available.

Insecticide Required. Some translators contain bugs. Sometimes these are poorly written, but often the problem is not being able to keep up with the current versions of GIS software.

Many argue that the GIS companies are not interested in being able to write out a bug-free file. Sure, they can read everybody else's file, but it is not to their advantage to be able to share data from their installations. They would like to be the center of the universe.

I'm not sure you can get there from here. Circuitous routes take time and energy. Sometimes the direct link is not possible because of missing translators (Langford 1987). There is not yet general agreement on what format should be used as an intermediate file transfer standard, so three or more steps may be required. I am hopeful that the new Spatia Data Transfer Standard (SDTS) will solve this problem for everyone.
If we both count with our shoes off we can get the same answer. Mixing single and double precision is asking for trouble. We found several cases where one GIS package worked in single precision, but the other was double precision (Craig et. al 1991; Bretano 1988). In this slide, points only 25 meters apart in double precision translate as being separated by 1000 meters in single precision.

Please use words that I understand. Complex primitives can be unintelligible. Universe 1 has no concept of a star. The IGES translation "standard" would try to push it through anyway—which never worked. A standard using simple primitives, like DXF or SDTS, would pass the star as a 10 sided polygon without problems.

Excessive File Size

As I get older, I find that eating such a large dinner can cause a sleepless night—and you can’t ask for a doggie-bag to bring home the leftovers. . . . It’s easy to get overwhelmed by the size of the file coming in. ASCII files waste a lot of space. Too big a file won’t fit on a floppy disk, even if it is compressed. Langford’s telephone bill in Alberta was big, because the data were a huge Landsat raster file. Stream mode digitizing can generate such files and it might be better to generalize such data to map scale—before storing them, and before attempting to transfer them to other users.

Vot’s a vertex is nodda node. A specific problem we encountered was the indiscriminate use of nodes and vertices when digitizing. (One translator we used converted everything to nodes or pseudo-nodes.) Because nodes are part of the topology of a database, more space is reserved for information about what is happening at each node. Appropriate use of vertices would reduce the file size.
Incompleteness

If regular clients are one-armed and one-legged, why go to the bother of producing full and complete business suits? Too many GIS operations cut corners that do not affect their own day-to-day needs—ignoring the norms of the real world.

Are we there yet? Sometimes the incompleteness is simply a matter of priorities. The data providers have not yet completed building their data set. But because of this, projects have been delayed or less desirable data sets have been used (Langford 1987). Good communication, truth in advertising, and flexibility in rearranging digitizing schedules could ameliorate these problems.

What goes in doesn't always come out. This is a very typical problem. Two lines are coincident in the real world. Because the major product of the home system will be a paper map, only the dominant line is digitized. That means that the subordinate data layer is incomplete.

"Mr. Wolf, how do I get to Granny's house?" Here one road passes under another, so the person digitizing interrupted the north-south road. A GIS program attempting to use this file to route an ambulance would by-pass this road, because it will appear as impassible. Humans would know how to read this map, but we're not so good at identifying optimal routes.

Geographic Positioning

You can't overlay geographic data files if you can't register your maps. And isn't map overlay one of the fundamental benefits of GIS technology? Lack of geographic accuracy on the incoming file can lead to spurious polygons and disconnecting lines.
Let me show you my etchings. We found some people using their GIS software to make schematic drawings which had no reference to geographic coordinates. This slide shows the distortion that can occur with such drawings. The real world block might not be square either, but I can't trust or locate any of the data that come from this provider.

Precision ≠ Accuracy. It's not uncommon for a city or county to digitize subdivision or tax parcel maps using very high standards. The problem is that the original map was of low geographic quality. In St. Paul, a newer file based on GPS technology was combined with a tax parcel map and the results had peoples' homes moved into the roadway.

Give me that address again. Undocumented coordinate systems can cause problems. These two points have the same coordinates, but one is referenced to the North American Datum of 1927—NAD27—while the other is based on NAD83. Programs exist to make the conversion, but it is critical to know the basis of the incoming map.

The natural world contains very few right angles. Data captured in raster format causes particular problems. The upper map on the right shows the result of a precise translation to vector. The lower map is probably closer to the real world, but who knows?

Lack of Topology
In the topological vector world, lines run continuously, polygons close, and a lake holds its water. Snapping is the digitizing technique that forces this connectivity: lines are connected mathematically, not just visually. When a CAD product is the goal, snapping is not so critical and often overlooked. Langford (1987) received a file that contained so many bad polygons, that he abandoned translation and redigitized the maps himself.
Blackouts. Have you ever used MacPaint, tried to shade a polygon that you didn’t quite close, then watched that pattern spread to fill the entire screen? MacPaint doesn’t have a snapping option—fortunately, it does have an “UNDO.” Too many files we received had polygons that didn’t close.

Spaghetti makes bad pavement. This is the linear version of not snapping to closure. One street overshoots leaving a nonexistent stub. The other never reaches the highway and frustrates both transportation modelling and on-board vehicle navigation systems.

Sheet faults. Lack of closure can often result from map tiles that have not been edge-matched. We had a conscientious contractor who mathematically connected his maps at the corners, but didn’t have the software to make certain that polylines ran continuously across the map borders.

Sometimes, things actually line-up in the real world. It is important to snap between layers, especially between coincident layers. In North America, the Public Land Survey is the basis both for land ownership and road networks. These two themes need to be snapped together.

Label Problems

Much GIS power comes from having data attached to map objects. The relationship between the map object and the data file is forged through an object ID or label. If the object has no label, which is reasonable in a CAD environment, a GIS cannot operate on it. It is just as bad to have two labels. I sure would like to know whether that one can contains popcorn or poison.

Double labeling can result from adjacent polygons not closing and the two individual labels being applied to the unwanted super-polygon. ARC/INFO attaches labels by placement and a large label on a small polygon could appear outside that area, causing problems for two polygons.
Identical or fraternal Twins? DXF passes an object signature as a unique combination of layer, line type, and line color (Whipple, 1989). If care is not taken to discriminate between object types during data capture, there is no sure way to know where to sleep at night and where to spend Sunday mornings.

References

Will Craig is associate director of the Center for Urban and Regional Affairs, University of Minnesota. He was president of URISA, 1986–87. His research interests include overcoming the barriers to sharing data with other agencies and the public. Craig holds a Ph.D. in geography.

The 5th Wave
By Rich Tennant

“YOU KNOW THAT GUY WHO BOUGHT ALL THAT SOFTWARE? HIS CHECK HAS A WARRANTY THAT SAYS IT’S TENDERED AS IS AND HAS NO FITNESS FOR ANY PARTICULAR PURPOSE INCLUDING, BUT NOT LIMITED TO, CASHING.”
At the 1990 URISA Conference in Edmonton, Alberta’s Land-Related Information Systems (LRIS) Project team presented the concepts for the planned LRIS Network and associated business entity. For readers new to Alberta’s LRIS Project, a brief history follows.

For almost two decades Albertans from the private and public sectors have been working towards the implementation of the LRIS Network. The province of Alberta has an area of 255,285 sq. mi., with 62 percent of the surface land and 81 percent of the subsurface rights (e.g., oil and gas) owned by the Crown. Automated systems for land-related information were deemed important tools for the management of these public resources, and have been developed and utilized over the past 10 years. The LRIS Network was originally a concept for sharing and exchanging information. It is now a technical implementation that initially links together a few databases, but that in time will have as many databases connected as there are suppliers of data wishing to participate. More detail on the technology of the LRIS Network is found later in this article. The cost of construction of the systems and mapping programs that will be included in Phase I of the LRIS Network development will exceed $100 million CDN.

As the costs to complete Phase I of Alberta’s LRIS Network were finalized, a strategy for operating the system and, in particular, offering its products and services to the public became necessary. The private and public supporters jointly proposed to the provincial government that the need to operate in a business-like manner was essential, and that the business should also be responsible for recovery of the final construction costs (of the LRIS Gateway and LRIS Spatial Database System), as well as operating costs. In December 1990, the year of the Edmonton URISA Conference, the Alberta government approved funding for construction of the LRIS Network. In September 1991, approval permitting the creation of the business entity—“Land Information Alberta”—was given. In this article we will describe the business and technical decisions that have been made and the issues that have emerged, as Land Information Alberta (LIA) opens for business.

The first phase of the LRIS Network is being implemented in three releases. The first, available commercially in Fall 1992, provides remote electronic access to textual data from freehold and Crown land registries, and a selected survey plan information database. The second release will incorporate a simple data integration process as well as new information products, and the third will provide user-defined data integration and an increasing array of data products. A spatial database and the functionality to integrate spatial and textual data is also included in Release 3. The technical component is discussed later in this article.

The Business

We have spoken of “going into business.” What does that mean? It is the intent of Land Information Alberta (LIA) to provide remote electronic access to, and integration of, land-related information on a cost-recovery basis. LIA’s mission statement: To deliver high-quality land information products and services to Albertans for their economic, environmental and social benefit. For an information-brokering enterprise, that mission translates easily into a corporate philosophy which accentuates the commercial exchange. This is appropriate for a business and is understood by our private sector advisors. There is, however, growing concern on the part of data suppliers that access to government data might eventually be constrained on the basis of business priorities, rather than public policy. Coupled with that notion is the ongoing concern on the part of the user
community that LIA could exercise its monopolistic advantages in the short term. The undesired consequences to customers would be accelerated debt repayment (to the provincial government through surplus revenues) resulting in high prices for access, and the products of data integration. The intention of “going into business” is to recover operating costs, repay debt, and finance enhancements to the products and services; however, the intent of “going into business” is not universally understood or accepted. LIA’s first business priority in the marketplace must be to clearly demonstrate its commitment to the production of quality products and services, at prices which are acceptable to consumers and appropriate to
sustaining business operations. LIA's resulting pricing strategy stresses cost recovery, and revenues based on sale volume rather than high prices.

LIA is being created as a unique business entity within the Alberta government. To maintain the basic business focus, an operating line of credit has been secured from the Provincial Treasurer. Without voted funding, the ability to maintain and enhance the products, services and organizational infrastructure rests entirely on LIA's success in the marketplace. Aside from construction activities, it's no surprise that the major focus has been on defining and creating the marketing unit. Within LIA, that general function will be part of what is called the Customer Relations Centre. Definition of the role of this unit was part of the "Market Entry Strategy," completed for LIA in December 1991 by a management consulting firm. This unit will be responsible for market development activities, communications, customer support through the provision of help-desk and user consultation services, and new product development. For opening day, LIA will have two market development officers, a marketing/communications officer and four customer service officers. All have been recruited (or seconded) and in training since mid-February. The product development group is still being defined. Until the 1993/94 fiscal year, new product development will be managed jointly by the LRIS Project team and LIA. The business administration function is the task for a second unit, Corporate Services. This unit will provide for the financial accounting, business planning, network operations, and general administrative needs of the organization. To date, this unit has a senior manager, an accounting supervisor and an administrative clerk. Most of the first year support to maintain the LRIS Network hardware and software applications has been contracted to specialist agencies within the Alberta government.

LIA received approval to staff 12 permanent positions. Those which have not been filled will be allocated on the basis of need. First priority would be two additional marketing and customer service positions; next are the positions responsible for data administration and network operations. Once these have been filled, and business activities warrant, an application will be made to obtain additional permanent positions.

Opening day draws closer as a concept takes form. Since 1987, four major business-oriented consulting reports have been completed. The market feasibility and marketing strategy have been well researched. The basic message in each case has been the confirmation that the LRIS concept is financially viable and that the demand for its products and services now exists. On that foundation, Land Information Alberta enters the marketplace. LIA will position itself as the information broker. This basic service will initially involve remote electronic access to land information. Both LIA and its users will grow together.

The beginning emphasis is straight-forward:

- Promote the benefits of use, not the system capabilities,
- Reassure users about data accuracy, reliability and currency, and
- Remove barriers to entry.

Underlying this must be strong and consistent user support. This will take the form of on-site demonstrations, information on how to make better use of the system, user training, a customer service hot-line and extended hours of service.

Among the major hurdles to overcome are user awareness and excessive expectation. Seemingly contradictory, they underscore the dichotomy of the marketplace. One may be knowledgeable and have high expectations for the LRIS Network, or little is understood about it save for the notion of remote electronic access.

There is a market for LRIS products. Particular care must be taken in marketing and managing its delivery, however. The most optimistic financial estimate does not predict a positive cashflow for at least 2–3 years from start-up. The potential is immense as the builders and users come to appreciate and expand the functions, benefits and usefulness of Alberta's LRIS Network.

The Technology

In the past year, Alberta has moved quickly in the implementation of its LRIS Network (Figure 2). Phase I of this development is expected to be completed in mid-1993. There are two major systems development
efforts underway as part of Phase I—the LRIS Gateway System, to be delivered in three releases, and the LRIS Spatial Database (SDB) System. Release 1 of the LRIS Gateway System will be available to the public in the Fall of 1992.

The basic components of the LRIS Network are the Primary Systems, LRIS Gateway System, and Thematic Systems. The Primary Systems are those systems that were identified as providing fundamental, common denominator land-related information. The completion of the Primary Systems and construction of the LRIS Gateway and LRIS SDB Systems, will complete Phase I of the LRIS Network.

Thematic Systems capture the remainder of the land-related information systems in the LRIS Network concept, and are generally described under the categories of Environmental/Natural Resources, Infrastructure, and Socio-Economic. While the Primary Systems are almost exclusively systems and data held by the Alberta government, the Thematic Systems will come from the private sector, municipal and federal governments, as well as the provincial government. The addition of Thematic Systems to the LRIS Network is Phase II of the LRIS Network development, and is expected to continue for many years.

The LRIS Gateway System—Release 1 is based on existing information products, from single systems (Figure 3). Release 2 adds new functionality and new products and services, such as different utilization of the systems (new keys, different fields retrieved, changed limitations on searches). New functionality will include the first integrated data products. These include products where the results of a query to one system are used in querying a second supplier system, or where two or more systems are queried, and the results are integrated into one product within the LRIS Gate-
way System before pick up. Providing customers with more flexibility in building queries will be important.

Release 3 plans for the availability of the LRIS SDB System in conjunction with the LRIS Gateway System (Figure 4). With this release the spatial information will be combined with the more traditional text information. Geographic/spatial and attribute driven queries will be supported. Release 3 will also support ad hoc query capabilities, within the technical limitations imposed by the supplier systems. Release 3 is planned for completion in 1993.

The LRIS Spatial Database System is a fundamental component of a complete and functional LRIS Network. Based on information from the 1:1000/1:5000 Municipal/Parcel Mapping programs, and from the 1:20,000 Provincial Digital Base-Mapping product, the LRIS SDB System will be a storehouse for that data. The LRIS Network requires that component elements such as survey parcel, road, and river be stored and maintained as unique, retrievable entities. The LRIS SDB System will provide the functionality required to do retrievals using spatial operators. The process of selecting a contractor to implement the system was completed in July 1992.

There are five underlying principles that have been adopted during the development of the LRIS Network:

1) That data stay with the supplier or agency with the mandate and responsibility for maintenance.

2) That customers from anywhere in the province can access the LRIS Network.

3) The system be an on-line service.

4) A heterogeneous hardware and software environment be supported.

5) That the investment in data not be compromised by the technology.

The LRIS Gateway System design has incorporated these principles.

The LRIS Network simulates on-line service. The user has a run-time copy of the LRIS Gateway System software on their computer. Connection with the LRIS Gateway System's VAX 4000 is made via modem and a commercial data carrier. The customer fills out the product order on his or her own machine, and sends the request to the LRIS Gateway System. Custom interfaces to the supplier systems execute the order and return the results to the LRIS Gateway System, where the data are loaded into tables. At that point the customer can pick up the results. Pick-up can occur at any time up to three days after completion. This two-stage process also provides additional security to the supplier systems, as it is only the LRIS Gateway System that directly accesses the supplier, and not external users. The source of data remains the supplier's database system.

The open nature of the software design will allow us to have different groups build their own custom interface to the LRIS Gateway System. This benefits local system developers/land professionals. The
open design is also expected to facilitate the development of business relationships with value-added retailers—businesses that take the LRIS Gateway System’s products, perform extra analysis or add other data, and re-sell. Finally, while the technical solution is a made-in-Alberta/for-Alberta solution, it has significant transportability by virtue of its design.

The technical approach has been to keep the LRIS Gateway System as open as possible, to allow for future migration to other technology and, more importantly, to allow for business to build on the system. The easy access to standardized information, availability of specifications for the LRIS Gateway System, combined with complete provincial data coverage, provide an excellent environment for the development of specialist software and businesses.

Summary

This article provided an overview of the business and technical considerations that influenced both Land Information Alberta and the LRIS Network as they prepare to open for business. Part II, in a subsequent issue of this journal, will report on how well the business and the technology are meeting expectations and demands of the user community.

Pat Merrick has worked on the Alberta LRIS Project since 1990. He has been responsible for defining and creating the LRIS business entity, “Land Information Alberta” (LIA). Until staffing is completed, he will continue as director of LIA’s corporate services. He has 20 years of managerial experience in the private and public sectors, primarily in the delivery of financial and administrative services, including the establishment of the Maps Alberta business entity.

Laurel McKay has been directly involved with the LRIS Network Project since 1983. She is director of the LRIS Network development as well as being involved with the development of Land Information Alberta. She has been a URISA board member since 1989, and is president-elect and conference chair for 1993.
Horwood’s Short Laws

Stephen Kinzy

In 1984 the late Dr. Edgar Horwood, founder and first president of the Urban and Regional Information Systems Association (URISA), retired from the University of Washington. In honor of his services to URISA and his contributions to the field of urban and regional information systems, a number of his friends and colleagues decided to give him a retirement dinner during the 1984 Seattle URISA Conference. The ’84 conference was the Association’s 22nd annual meeting. The theme of the conference that year was “The Changing Role of Computers in Public Agencies.”

During the summer of 1984, prior to the URISA Conference, Dr. Charlie Barb, Peirce Eichelberger and I were working for HDR Systems in Omaha, Nebraska. Charlie had been one of Ed’s Ph.D. students at the University of Washington and a long-time friend. Peirce and I became acquainted with Ed through our GIS consulting business and URISA Board activities. Prior to the 1984 URISA Conference, the three of us discussed preparing a commemoration for Ed’s retirement dinner. We agonized over several options. Then I recalled a humorous list of information systems laws that Ed had given me during the 1979 URISA Conference in San Diego.

I remember well the circumstances when I received the “Laws.” One of the recreational activities during the 1979 conference was a boat tour of San Diego harbor. Rosemary Horwood, Ed’s wife, was very active in coastal land use planning and management at the time so she and Ed were on the tour. During the tour Ed and I talked about the difficulty of convincing public officials to make a commitment to information systems. Ed suggested that a little humor might help get the message across. At that point Ed dug into his wallet and gave me a crumpled piece of paper entitled “Horwood’s Short Laws of Data Processing and Information Systems.” He said that the laws had been developed on the basis of his experience over a number of years. I don’t know when they were originally created or if a longer version of the “Short Laws” ever existed, but they are appropriate and certainly timeless. I eventually had a set of slides made of Horwood’s Laws, and over the years have used them many times in GIS seminars I have conducted. Audiences seem to enjoy the humor and remember the wisdom.

So as the 1984 conference approached, Charlie, Peirce and I were struggling with an appropriate gift for Ed’s retirement dinner. We eventually settled on a poster of Horwood’s Short Laws. Charlie contacted Rosemary without Ed’s knowledge, and she sent us a picture of Ed which we incorporated into the poster. Charlie, a long-time URISA member worked with Rolf Scamitt, an equally long-time URISA member, to compile a list of all URISA conferences and officers from 1963 until the 1984 meeting in Seattle—an enormous effort if you knew the state of the URISA records at the time. This list was printed on the back of the poster. HDR Systems funded the project and we printed a thousand posters for attendees at the 1984 conference. We also had a plaque made of the poster for Ed’s retirement dinner. The dinner was held during the conference; nearly 100 students, friends and family attended. It was one of the most memorable events of my life. It was moving to hear the tribute and adulation from students and friends all over the world that Ed had befriended and influenced. During the dinner, Peirce and I presented Ed with the plaque.

When Charlie, Peirce and I decided to put the poster together, we originally conceived the project as a tribute to Ed’s wit and humor. However, when Charlie and Rolf tabulated the list of URISA conferences and officers, the poster represented the legacy of a man dedicated to the advancement of technology for public good. That legacy
is still alive and strong, Ed would be very proud. At the bottom of the poster Peirce and I wrote the following:

In honor of Dr. Edgar M. Horwood's retirement and his contributions to the field of Urban and Regional Information Systems, and to URISA which he founded and through which his work will be carried on for the betterment of mankind and our spaceship earth. Thanks Ed, for a job well done. May the rest of us, who share your dream and vision for the future, keep the faith...

**Horwood's Short Laws**

1. Good data is the data you already have.
2. Bad data drives out good.
3. The data you have for the present crisis was collected to relate to the previous one.
4. The respectability of existing data grows with elapsed time and distance from the data source to the investigator.
5. Data can be moved from one office to another, but it cannot be created or destroyed.
6. If you have the right data you have the wrong problem, and vice versa.
7. The important thing is not what you do but how you measure it.
8. In complex systems there is no relationship between information gathered and the decision made.
9. Acquisition of knowledge from experience is an exception.
10. Knowledge grows at half the rate at which academic courses proliferate.

---

Stephen Kinzy is regional manager for ESRI in St. Louis, MO. He has more than 20 years of experience in GIS with the city of Omaha, HDR Systems, and McDonnell Douglas. He served on the URISA Board of Directors from 1978 to 1982, and is a recipient of the "URISA Leadership Award."
EDITORIAL INTENT

The reviews section of the URISA Journal provides critical reviews and information concerning publications, information sources, and software in the field of urban and regional information systems. Professionals in this field must draw from a variety of sources and materials to support their activities. Accordingly, this section includes reviews of a variety of information sources that, in general, fall into one of four broad categories:

1. **Book reviews** focus on books relating to the development, implementation, application, evaluation and management of information systems in urban and regional environments. Books written for use in academic settings, as well as those aimed at the practitioner, will be included.

2. **Publication reviews** examine a variety of publications that offer significant contributions to URISA’s areas of interest. These will include conference and symposium proceedings, reports, resource papers, monographs, technical publications, resource directories, workbooks and development guides and manuals related to urban and regional information systems issues.

3. **Video reviews** provide another review category. Videotapes are becoming a popular tool for education and information dissemination, particularly in the area of geographic information systems. The URISA Journal will include reviews of new videos that could be useful information sources or tools for Journal readers.

4. **Software reviews** are intended to help readers become aware of software that could aid them in their work, and to help them choose the right software by reviewing it from a URISA professional’s viewpoint.

As with all reviews, software reviews express the opinions of the authors. The publishers of commercial software will be given the opportunity to check review drafts for factual errors, but cannot change opinions expressed in the reviews. URISA Journal does not encourage complimentary submissions of software for review. We prefer that software publishers identify expert users of their software and then encourage them to write a review. Potential reviewers are encouraged to evaluate programs they use frequently, decide whether those programs would be of interest to Journal readers, and find cut if the programs have been adequately reviewed elsewhere. Many programs useful for our professions are not reviewed, or are reviewed by journalists, rather than by members of our professions.

The software reviews are categorized as follows:

- **In-Depth**: Detailed reviews of substantial programs. An In-Depth Review covers areas including ease of use, performance, and support. A program summary and overall rating complete the review.
- **Head-to-Head**: Program comparisons that take a class of programs such as thematic mapping software, introduce each package, compare their strengths and weaknesses, and discuss their use.
- **From the Inside**: Explanations of non-commercial programs by their creators, to inform readers about program capabilities and encourage use of the programs.

In addition to submissions of book, publication, video and software reviews, we welcome suggestions and insights pertaining to the identification, critical review, and recommendations of all information sources in the field of urban and regional information systems.

Rebecca Somers
Peter Van Demark
Book Review

An Introduction to Urban Geographic Information Systems

Reviewed by Kenneth J. Dueker


Editor’s Note: After going so long with so few GIS books available, we are now faced with a virtual explosion of texts, references, and related books on GIS. These books deal with various views and applications of GIS, reflecting the diverse character of the GIS marketplace. To help our readers choose from among the available books, it can be useful to provide reviews of books based on different perspectives.

The following book review concerns a GIS book recently reviewed in the Journal (Allan H. Scamidt, Vol. 3, No. 2). This review, provided by Kenneth J. Dueker, offers a different perspective on William Huxhold’s book. It is reprinted here with permission from the Journal of Planning Education and Research.

Recently, there have been a number of textbooks dealing with the application of GIS concepts and technology (Antenucci 1991; Aronoff 1989; Clarke 1990; Pequet and Marble 1990; Star and Estes 1990; Tomlin 1990). Huxhold’s book differs in two respects; it has a strong urban focus and it is rich in examples drawn from Huxhold’s experience in municipal government. Yet, all of these books suffer from the need to be introductory about GIS concepts and technology, while at the same time attempting to describe a wide range of complex applications. Huxhold’s focus on municipal applications, particularly housing and community development, minimizes the range problem. His good use of examples helps to teach the concepts and technology, while also illustrating in detail a number of GIS applications in municipal governments.

Huxhold starts with “Information in the Organization.” His example of the shift from transaction-based tax assessment to a data-based property records system illustrates well the benefits of applications programming in a database management systems environment. He also develops “The Urban Information Pyramid” that illustrates, with examples, the importance of horizontal and vertical integration of data in municipalities.

There is a detailed treatment of the different methods to store attribute data—flat files, hierarchical files, network structures, and relational databases. Again, his use of examples helps to distinguish among these database management approaches. Good use of examples also helps to illustrate the concept of topology and its importance in GIS.

The chapter on applications describes using GIS to address various issues. However, the issues are not related nor was a framework provided to help assess the applications back to GIS.

The chapter on geographic base files is particularly strong. It stresses the importance of a digital street centerline file for municipal applications and explains in detail the file structure of DIME and TIGER, prepared by the U.S. Bureau of the Census, in providing the ingredients for a geographic base file for local applications, such as address matching of administrative records.

The chapter “The Model Urban GIS Project” provides useful hints on evaluating geographic information needs, gaining organizational support, and managing the GIS project.

The county GIS problem set is an important contribution. It provides students an opportunity to create a database and work through a set of GIS applications, pertaining to topology and land records. An infrastructure module would be a useful addition, however.

In spite of its above described strengths, the book has four deficiencies in teaching “urban GIS.” First, the housing and community development in municipal government emphasis leaves untouched an important field to planners—small-area
data systems to support the urban transportation-land use planning process at the metropolitan level. Second, the book is surprisingly weak in the area of public works and utilities. Although Huxhold deals with cartographic data structures of GIS and AM/FM systems, he does not deal explicitly with applications in public works and utilities. This is a serious problem because in most municipalities the competition is between public works/Utilities and planning for GIS resources. The applications and accuracy issues that separate these forces need to be fully understood. Third, the lack of treatment of raster/vector integration does not open the door to applications of life cycle management of infrastructure, site planning, or urban change detection using remote sensing. Fourth, private sector GIS applications using small area and parcel data for spatial marketing and locational analysis are not addressed.

I am still undecided as to how to use Huxhold in our GIS sequence. At Portland State University, GIS is taught jointly by Geography and Urban Studies and the intro course covers both urban and natural resources, and vector and raster data structures. A follow-on applications course goes into more detailed examination of GIS applications areas, but again it covers a broader set of applications than are contained in Huxhold. I will use Huxhold’s book in the GIS Applications course along with one of the others, probably Aronoff. We will continue to use Star and Estes in the GIS Introduction course with selected material from Huxhold. Huxhold will become a recommended text for students who are focusing on urban applications.

Kenneth J. Dueker is the director of the Center for Urban Studies and professor of urban studies and planning at Portland State University. He is an editor of the URISA Journal.

References
EDITORIAL INTENT

A good picture is worth several thousand words, and so is a good map. Pages of data can be summarized in a single map, and spatial patterns become apparent that are impossible to discern among pages of numbers or words.

Computer mapping and geographic information systems play a key role in URISA’s mission to help local, regional and state/provincial governments make the best use of information system technologies. And now that computer-generated maps have become an important part of URISA’s domain, the URISA Journal showcases the efforts of mapmakers by featuring an exceptional map product in each issue. We use the word “map” in a very broad context and are willing to consider remotely sensed images and other graphics.

Our intent is to share a wide variety of maps, hoping to inspire others to learn and copy good ideas and techniques. We are looking for maps that are easy to read and that allow the reader to readily see what the mapmaker intended. Perhaps new symbols were designed or special colors chosen to meet a particular purpose.

New ways of presenting data or of understanding phenomena through mapping techniques are also desirable. And, of course, we will also consider publishing a map that is exceptionally pleasing or beautiful.

William J. Craig

Walker has solved this dilemma by providing both views on a single screen. In fact, he has provided many ways to view the landscape, including profiles and changing atmospheric and daylight conditions. His system is designed to be used in real time, so little emphasis is placed on hard copy maps. Consequently, our images are all made from photos of the computer screen. The purpose of these images is to convey an understanding of the world—as it really looks, or as it would look under different conditions. I think that Walker has shown the value of complementary views for conveying this understanding.
Applications for Digital Terrain Modeling

Robert S. Walker

Geographic Information Systems (GIS) are now an established branch of Information Technology. A GIS will integrate data by reference to its position on the surface of the earth. Usually these data are overlaid on a map, either on-screen or on hard copy, creating a two-dimensional representation of the data.

However, by using data representing the height of the ground surface above datum, a more sophisticated three-dimensional map can be produced. This data may be taken from existing contour maps, usually generated by ground survey, or obtained by remote sensing and photogrammetry to generate a set of spot heights. From this data a Digital Terrain Model (DTM), a mathematical model of the surface, can be created by the computer using a triangulation process to create the surface of the model.

By using this technique, slope and aspect maps can be created to analyze the terrain, or cross-sections created for line-of-sight calculations. The DTM can be visualized from a particular perspective view point, either on or above the surface, and displayed graphically on either a high-resolution workstation, or a raster plotter. Topographic features such as roads and rivers, or remotely sensed imagery, can then be draped over the terrain. Alternatively, solid objects such as buildings can be placed upon it. The result is a far more realistic impression than that given by the traditional flat surface map.

Applications of DTM

DTMs are used for the visualization and analysis of terrain; many of the early models were for military applications, which has remained an area of great importance. DTM provide the images for flight simulation, used for pilot training, and this has developed into ground-following systems for flying aircraft and missiles. Cross-country movement analyses use DTM for modeling vehicle and troop movements, incorporating data on ground surface cover such as soil type, hydrology and vegetation.

Increasingly, DTM are being used in other fields. Among these are applications in environmental assessment and radio propagation analysis.

Environmental Applications

DTMs are now being used to assess the effects of many types of influence upon the environment, from natural phenomena to man-made developments. For example, intervisibility maps created on-screen enable identification for the area within which a specific feature, such as a new highway or pylon, is visible. This offers the benefit of allowing the user to judge the impact a proposed development will have upon its surroundings, rather than having to rely on conjecture. GIS users are already using this facility for evaluating the visual impact of forestation schemes and proposed construction. The cross-section display facility enables line-of-sight questions to be answered.

The finished visualization can be presented in either “quick-see” wireframe format sketched on screen, or as a fully rendered, full-color visual with hidden surfaces removed. Other elements that could affect it, such as foliage, can then be added. The final view can then be generated from any angle, immediately highlighting potential problems. This also forms the basis for photo montage presentations, where photographs of buildings, for example, can be incorporated within the DTM.

In the UK, such presentations are now being used in public inquiries into potentially contentious development plans. In addition, the varying effect of natural light on the immediate environment can also be as-

Note: Because the page divisions, called signatures, fall in groups of eight, the feature map section falls out of sequence from its usual placement.
sessed, and effects such as low cloud or fog may be simulated.

DTMs are now being used for surface water run-off and reservoir studies, which are becoming increasingly important in the management and conservation of water supplies. Flooding and dam-break simulations are of similar value to contingency planning. Another future application currently in development is three-dimensional modeling of pollutant concentration, involving the integration of more complex fluid-dispersion models.

Radio Propagation Analysis

The DTM is now being used more and more in the field of international telecommunications. The key to building successful telecommunication systems is propagation modeling—the analysis of terrain and landscape objects to ensure that the signal strength received from a transmitter is sufficient for satisfactory communications.

Telecomm technology is rapidly expanding in all countries of the world. Land communication links include microwave transmissions for telephone networks, radio and television broadcast, emergency services and cellular telephone networks. For fixed point links using antennae, this analysis involves detailed propagation modeling. At the other extreme, cellular network planning has to consider elements such as the interference caused by radio frequencies in adjacent cells, particularly at boundaries of jurisdiction areas or following reorganization of frequency allocations.

Radio propagation planning needs three specific data sets to be effective and accurate—a terrain model, a database containing information on the effective height of ground cover such as buildings and trees ("clutter"), and a prediction algorithm.

When integrated into a GIS, the system allows total flexibility in predicting transmitter area coverage or producing point-to-point link profiles. It can also take the map data sets, including a terrain model, and allow the user to perform "what if" scenarios. For example, changing antenna height and type and varying the transmit power can all be simulated at the touch of a button, and results can be overlaid on existing map backgrounds.

Ease of Usage

Detailed GIS applications require large amounts of data, and involve complex processing. Consequently, a fully integrated set of GIS facilities for the maintenance and analysis of a geographic database is required. These functions need to be accessed through a standard set of easy-to-understand menu commands. Laser-Scan’s HORIZON GIS, for example, uses the industry standard OSF/MOTIF user interface, which minimizes the need to use the keyboard.

There is no doubt that the construction and analysis of DTMs will become an increasingly important aspect of GIS, as new applications are developed to solve new problems.

Dr. Walker has an M.A. in mathematics from Oxford University, and M.Sc. in fluid mechanics and Ph.D. in applied mathematics from Manchester University. He worked for more than ten years in the public utilities, and is currently marketing manager at Laser-Scan, the geographic systems software company based in Cambridge, England.
1b. Profile

The top image (1a) is a plan view of an area with color-coded elevations. Sight lines radiate from a transmission tower. Image 1b shows an elevation cross-section along one of those lines. In operation these two windows appear on one screen.

2b. Rendered View

Image 2a is another plan view. The other window (2b) is a rendered view showing how the scene would appear to a person looking south across the valley.

Rendered view from a DTM showing proposed ski runs. Winter coloration. Light shading to bring out terrain aspect.

Rendered view from a DTM with atmospheric effect (fog).
Dynamic Segmentation Revisited: A Milepoint Linear Data Model
Kenneth J. Dueker and Ric Vrana

Considerable attention is being given to "dynamic segmentation" as GIS vendors scramble to include it in their repertoire for marketing to transportation organizations. Surprisingly, there is little consensus on the concept, and implementation approaches vary among GIS products. There is some confusion as to what dynamic segmentation entails, and in understanding and evaluating the approaches offered by system vendors. A framework for comparative analysis is needed to evaluate the relative advantages of different approaches.

Perhaps it is timely to re-examine dynamic segmentation and restate the problem in fundamental terms to encourage consistent approaches. A comprehensive analysis of the problem would consist of the following efforts:

- Systematic analysis of the requirements for dynamic segmentation applications,
- Clarification of concepts, terminology, and data models,
- Examination of different vendor approaches and algorithms,
- Development of comparison criteria within a benefit/cost framework to evaluate the trade-offs inherent in different approaches, and
- Evaluation of products in a head-to-head comparison using the framework to assess tradeoffs associated with the different approaches.

This paper addresses the first two steps only, although examples from vendor approaches are used to illustrate the problems with existing terminology and differences in approach and implementation.

The requirements for dynamic segmentation reveal fundamental principles concerning attribute and spatial data handling for linear spatial objects. Toward clarifying these principles, a milepoint (MP) linear data model is presented to relate distance-referenced spatial and attribute data. This is followed by a discussion of distance referencing of attribute data and of the requirements for a dynamic segmentation algorithm. Several vendor approaches to dynamic segmentation are then described, but a more systematic evaluation of approaches is beyond the scope of this paper.

The Nature of the Problem

Dynamic segmentation is a particularly important issue in applying GIS concepts and technology to the field of transportation. The focus of this paper is on dynamic segmentation as used in highway organizations, though the principles are applicable to stream and utility applications as well.

A wide variety of data on incidents and conditions along centerlines and within rights-of-way are used to monitor the safety, maintain the infrastructure and analyze the conditions on the nation's highways. Incident data may be gathered on the location of traffic accidents, signage, or service facilities. Other data describe lengths of constant values of the roadway pavement, width, or jurisdiction. The occurrences of these incidents and lengths are characterized as point and linear features and can be described in databases with thematic and temporal attributes ("collision with pole," "at 1800hrs, March 2"). They are linked to location along a highway route by attributes that identify the route ("Highway 48") and a position along that route such as milepoint ("MP 13.12").

For many applications, such as pavement management or traffic inventories, certain sections of highway routes may have been predefined into sections. Data on roadway width or counts are linked to these sections, which can be mapped by associating their beginning and ending milepoint references to coordinate space. However, the a priori segmentation for one attribute often overlaps
with that for another, making it difficult to analyze their interaction and coincidence. Moreover, not all incidents or conditions of interest may be referenced to the same locational framework. Knowing the jurisdiction of a feature along a centerline may require associating the route milepoint of the feature to boundary data stored in state plane coordinates.

One solution is to predetermine every possible item of interest, relate all point and linear features or lengths to one reference system, and predefine sections for the myriad of unique combinations of all possible attributes. This is unwieldy. Alternatively, establishing small segments on an arbitrary basis, such as hundredth-miles, and storing all attributes for each unit provides a solution, but at the cost of managing many spatial units with the same sets of attributes.

A two-dimensional analogy to the problem occurs with many natural resource applications. Areas can be exhaustively partitioned into discrete and homogeneous polygons according to the presence of an attribute of interest. Polygon overlay is the procedure by which the interaction and coincidence of these attributes is managed. A map of one attribute (e.g., soils) can be related to a map of another (vegetation) to produce a composite. Each attribute, however, is maintained as a separate layer.

Relating attribute and spatial data seem to pose a greater problem with linear than for point or area features. The large degree of attribute overlap along linear features probably accounts for this. This results in greater difficulty in maintaining a one-to-one linkage of spatial and attribute data for linear features such as highways. Partitioning linear data into discrete and homogeneous segments depends on which attribute is selected to be held constant.

With dynamic segmentation, linear entities do not need to be decomposed into homogeneous segments, a priori. The road exists as a series of spatial objects and features along it are linked to linear spatial objects by distance referencing. Dueker (1987) identified run-time segmentation using distance referencing as a solution to the problem of overlapping attribute data. Meanwhile, users were beginning to suggest solutions and enhancements to vendor products (MacDuff 1987, and Fletcher 1987). This was characterized as dynamic segmentation (Nyerges and Dueker 1988) and has been elaborated by Nyerges (1990), wherein geographic features, such as highway routes, are modeled as linear spatial objects.

Distance referencing of linear data provides an opportunity to avoid the direct need to rely on line overlay of single attribute-defined objects (see Fletcher 1987 for a development of the line overlay approach), or the aggregation of attribute-homogeneous lines. Dynamic segmentation of linear spatial objects provides a means by which new point or line objects can be created by relating the distance-referenced attributes with a manageable set of distance-referenced linear objects. Dynamic segmentation removes the need for a set of spatial objects for each attribute. Spatial objects and distance referencing of routes are used to create attribute-based spatial objects as needed.

Two more considerations are important for understanding the nature of the problem of dynamic segmentation. Many dynamic segmentation analyses result in displays wherein the new linear spatial objects do not have to be saved. For other analyses, the new linear spatial objects may need to be saved for input into a line-in-polygon search, e.g., to determine the miles of highways traversing wooded land. In addition, saving the new linear objects facilitates pan and zoom, even though they can be recomputed with comparative ease in new workstation environments.

**Dynamic Segmentation Concepts**

What is commonly called dynamic segmentation addresses three data handling issues in a GIS. For the sake of convenience, we may label these issues Linkage, Segmentation, and Display and Spatial Analysis.

*Linkage* pertains to the linking of spatial and attribute data by relating records for linear spatial objects and for distance-referenced attribute data.

*Segmentation* results in the creation of new point or linear spatial objects at run or query time, based on the choice of attributes and values needed for a specific application.
Display and Spatial Analysis involves an explicit referencing to a two-dimensional spatial framework of the attribute-based linear objects resulting from the linkage and segmentation. This can be specified in terms of pixel-space for the display of objects on CRTs and coordinate space to store the newly formed objects if needed for subsequent line intersection, buffering, line-in-polygon, and other spatial analyses.

In dealing with these issues, the following criteria for dynamic segmentation are recommended. An implementation must provide:

- Distance referencing of both thematic attribute data and linear spatial objects,
- Distance-referencing resolution for interpolation,
- Searching by thematic attribute or spatial queries (See Figure 1),
- Highlighting of selected point or line features in displays,
- Tabular reporting of attributes of selected point and line features, and
- Storing of new linear spatial objects if needed for subsequent spatial analysis.

Types of Linear Data Used by Highway Organizations

Nyerges (1990) and Briggs and Chatfield (1987) identify three highway reference schemes:

- Road name and mile point
- Control section
- Link and node

The challenge is to relate these three schemes in a GIS-T. We redefine these highway reference schemes to types of geographic features as follows:

- Reference markers—physical objects along roads that may or may not have a simple relationship to the length of roads and that form control points with a route and MP measure;
- Sections—short segments of highways such as used in pavement management systems or in the Highway Performance Monitoring System (HPMS), or between reference markers;
- Routes—collections of sections wherein sections may belong to more than one route;
- Control sections—uniquely identified route or collection of sections, and
- Chains—topologic arcs or links in a network, relatable to sections.

The relationship among these features needs to be specified to create databases that describe the real world robustly, as well as abstractly and efficiently.

Linear Data Models

This section presents the data model on which dynamic segmentation is based. Explicit linear data models are necessary to compare approaches.

Prior to dynamic segmentation, thematically defined linear features were treated as entities and represented as spatial objects. Two data models could be used to represent linear data, one being a Computer-Aided Design and Drafting (CADD) linear data model and the other a Data Processing (DP) linear data model. In a CADD linear data model, Layers/Levels and Line type, weight, and color are
used to define linear spatial objects (See Figure 2).

Overlay of line layers can be used to provide visual inspection of intersecting layers/levels, and line types of interest. If milepoints are maintained as attributes of the spatial objects, the results of line overlay can be saved as milestone limits.

Alternatively, the linear features can be sliced into small spatial objects that are homogeneous with respect to attributes. Here the entity is a small, uniform-length spatial object, referred to as the DP linear data model. In the DP linear data model, what were layers and line types in the CADD model are now attributes and values (See Figure 3).

Neither the CADD nor DP linear data models have proved to be satisfactory data models for use in GIS. Both require too many spatial objects to manage. A more robust linear data model is possible, one that employs distance referencing, defined as a one-dimensional measure along linear features.

Models of geographic data systematically relate data records describing the spatial characteristics, i.e., the location of a feature, with data records describing thematic attributes, i.e., the type of feature. In the simplest case, a model to represent this is a one-to-one relationship between the thematic attributes of a named or numbered route and a linear spatial object defined by a string of coordinates, i.e., a one-dimensional spatial object. This is called a polyline in Spatial Data Transfer Standard (SDTS) terminology (USGS, 1990) (See Figure 4).

---

**FIGURE 2.**
The CADD Linear Data Model

---

**TABLE:**

<table>
<thead>
<tr>
<th>Layer/Level</th>
<th>Line Type, Weight, Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanes</td>
<td>2 4 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Functional Class</th>
<th>Primary</th>
<th>Secondary</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Poor</th>
<th>Good</th>
</tr>
</thead>
</table>
The necessity of representing the entire route by a single linear spatial object can be relaxed by allowing a one-to-many relationship. This facilitates flexibility in terms of capturing or digitizing the spatial data, but still does not allow attributes to vary within the set of spatial objects (See Figure 5).

An additional problem is that the relationship of the one-route to-many-spatial-objects relationship only specifies a set of spatial objects that make up a route. The correct order or sequence of spatial objects is not specified, nor is there a requirement that the spatial objects connect to provide a continuous route. The sequence and direction of spatial objects can be specified by a distance-referencing relation.

A distance-referenced route is defined by a measure along it from an arbitrary starting point. For highway applications, this is usually done by identifying point and linear features to the nearest 0.01 mile. Important features such as intersections, bridge abutments, jurisdictional boundaries, beginning of projects, etc. are identified by reference markers, from which field measurements can be made. Two reference markers are needed for each linear spatial object to sequence them and give them a direction. A route can have many reference markers, as can a linear spatial object.

A known coordinate value for a reference marker can be inserted into the coordinate string of its spatial object. The more likely case is to calculate the x,y coordinate value using the coordinate string values and the milestone of the beginning of the string, or the location of another reference marker. Then the computed value is inserted in the coordinate string of the polyline for the linear spatial object.

This model also provides for a many-to-one relation between a route and linear spatial objects, which allows routes to overlap. This is necessary because routes may converge and share the same roadway for some distance. Similarly, transit routes can converge and share the same arterials in downtown areas. The data model should be robust enough to handle these situations, even though some states employ control sections or uniquely numbered state highways to maintain the one-to-many relationship of highway section to spatial objects. Forcing all state DOTs or all transportation applications to do this is overly confining (See Figure 6).

Sections (linear features) can be used rather than reference markers (point features) to relate routes to linear spatial objects. A route is an ordered set of sections, and sections relate to an underlying set of linear
spatial objects. For each Route-Section to linear spatial object relation (table) a corresponding From-MP and To-MP is required, although vendors can relax this kind of restraint (See Figure 7).

This data model is robust enough to handle inventory requirements. It is still deficient with respect to routing applications, though. This may be important when extracting weight or width restrictions from the highway inventory data stored by milepoint or sections, such as routing hazardous materials or overweight/oversize vehicles.

To handle routing applications, a topological link relation is added that allows building paths in a network (see Figure 8). Combining the linear spatial object and link, in a one-to-one relation, produces a network chain (in SDTS terminology). Figure 8 shows that a linear spatial object relation is required, but if topology is added, the linear spatial object polyline and topological link form a network chain. Some GIS architectures support dynamic segmentation on topologically structured data, while others handle dynamic segmentation in a module that is independent of topologically structured data. Those that have topological data structures have a required one-to-one relation between the link and the spatial object in order
to form the chain replacing the direct linear spatial object and route-section relation. GIS architectures that do not require topology must allow an optional link for transfer of data to another module for routing and network analysis.

This set of relations constitutes the Generalized Milepoint (MP) Linear Data Model. This is the model that can be used to compare vendor offerings of dynamic segmentation.

With the MP linear data model, finding the location of a given Route-MP value becomes a process of interpolation or proration along a coordinate string, using the From/To milepoint values. This is analogous to the urban geocoding problem for street address matching, wherein a specific address is interpolated using address ranges of street centerline files, such as TIGER/Line. In fact, reducing it to a geocoding problem may make existing code usable by some vendors.

More on Distance Referencing of Attribute Data

Data relating to linear features, such as roads, can be characterized as thematic entities located by measuring from known points, e.g., reference markers. There are two types of entities:

- Point features—accidents, signs, culverts, etc.
- Linear features—a length of road having a homogeneous characteristic, such as a constant pavement width, condition, or surface type.
Linear features are referenced by From-MP and To-MP. A point feature can be attributed with one MP. Databases containing the feature ID and the MP references are called route or milepoint attribute tables.

Features can be structured in two ways in route or milepoint attribute tables. A common approach in road inventory systems is an integrated database containing a number of different types of data in a single database. Locational attributes of these features are expressed as route name and MP. This integrated database approach facilitates the identification of new linear spatial objects using multi-attribute structured queries (these might also be called "thematic objects given spatial expression as linear spatial objects (polyline)"). The From-MP and To-MP result of querying the multi-attribute database is passed back to the graphic processor for highlighting the length of road meeting the search criteria.

A more frequently used approach is to have separate route or milepoint attribute tables for each data type, i.e., signs, culverts, and accidents. Again, each feature is located by milepoint referencing. Complex multi-attribute queries can be structured using logical combinations of attributes from the separate milepoint attribute tables, resulting in the From-MP and To-MP of highway features meeting the search criteria.

A robust dynamic segmentation implementation in a GIS must be able to handle both integrated and separate forms of attribute tables for point and linear features, and dynamically segment them to create spatial objects for display and/or spatial analysis.

Dynamic Segmentation Algorithm Requirements

The MF linear data model and distance referencing of attribute data are combined in a dynamic segmentation algorithm to perform the following:

- Create linear spatial objects from a Boolean combination of attributes in route or milepoint attribute tables,
- Display a symbol for a point feature at a milepoint location on a linear spatial object,
- Highlight a linear feature between milepoint locations that may be a linear spatial object, a part of one, or combination of several,
- Display linear spatial objects by interpolating from the x,y coordinate strings or by interpolating the pixel representation of them, and
- Save the new linear spatial ob-
Problem: IF  
Pavement width = 26' begins at MP\textsubscript{B} and ends at MP\textsubscript{D}  
Find \(X_\text{B}Y_\text{B}\) and \(X_\text{D}Y_\text{D}\) for analysis  
and highlight line segments between them for display.

GIVEN:  
\(X_\text{A}Y_\text{A}\)

\(X_\text{C}Y_\text{C}\)

\(X_\text{E}Y_\text{E}\)

\(MP_\text{A}\)

\(MP_\text{B}\)

\(MP_\text{C}\)

\(MP_\text{D}\)

\(MP_\text{E}\)

And \(X_iY_i\) for all shape points.

Step 1:  
Obtain MPs and chain(s) for beginning and end of pavement width = 26'.  
SQL query to database returns From\_MP, Chain\_ID  
To\_MP, Chain\_ID  
Where pavement_width = 26'

Step 2:  
Obtain proportional offset for From\_MP and To\_MP with respect to Chains.  
Compute ratios \(AB/AC, CD/CE\)

Current Approaches to Implementation of Dynamic Segmentation

Four vendor approaches to dynamic segmentation are reviewed, but not in detail as the software has not been exercised. Four vendors are currently marketing dynamic segmentation to state DOTs. They are:

- ARC/INFO™, distributed by the Environmental Systems Research Institute (ESRI), Inc. (Redlands, CA, (714) 793-2853),
- TransCAD™, distributed by the Caliper Corporation, (Newton,
FIGURE 10.
Dynamic Segmentation: Display

For Vector Analysis
Step3:
Using coordinate geometry and shape point XY, increment along successive line segments on each chain containing a From_MP or To_MP until proportional offset is exceeded.
Step 4:
Compute MP for last shape point before proportional offset is exceeded.
Step 5:
Interpolation along line segment to locate B at From_MP and D at To_MP.

For Raster Display
Step3:
Count total pixels along Chains AC and CE.
Step 4:
Apportion proportional offset B along AC and D along CE by counting pixels.
Step 5:
Highlight individual pixels while condition Proportional offset < incremented count is True.

MA, (617) 527-4700,
- MGE™/Segment Manager and MGE/Dynamo, distributed by Intergraph Corporation, (Huntsville, AL, (205) 730-2000), and
- GDS™/DSIM, distributed by Electronic Data Systems (EDS) (Maryland Heights, MO, (800) 437-4491), (formerly marketed by McDonnell-Douglas).

All vendors employ essentially the same data model, with some differences in implementation. TransCAD, ARC/INFO, and Dynamo require topologically structured data, chains (links and linear spatial objects). This facilitates routing applications of the dynamic segmentation results without subsequent restructuring of data as is required in GDS/DSIM (Dynamic Segmentation, Incident Management) and MGE/Segment Manager.

In GDS/DSIM and MGE/Segment Manager the milepoint linear data model does not require topology, although the results of a dynamic segmentation session can be converted to a topological data structure in another module (PCN for GDS/DSIM and MGA for MGE/Segment Manager) for subsequent routing analysis. ARC/INFO and TransCAD will accept long cartographic strings, but they break the strings into chains and build topology.

Although users may not see sections split at topological nodes, a vendor may do it to maintain the one chain to many sections relation. ARC/INFO has route-section tables while GDS/DSIM and MGE/Segment Manager provide separate fields in the linear spatial object record to handle multiple routes that share common sections. In TransCAD, chains that are common to different routes are duplicated. Users with extensive overlapping routes may benefit from route-section tables, while users without much overlap in routes may find route-section tables just more data to manage.

TransCAD utilizes an approach that interpolates between endpoints of control sections to assign milepoints to endpoints of chains. The TransCAD linear data model consists of uniquely numbered highways called control sections that are a collection of chains, and chains are a collection of
sections. Dynamic segmentation in TransCAD is essentially a geographic problem, interpolating along chains using milepoint ranges. TransCAD creates and saves new spatial objects whenever a value changes for any of the query attributes. This facilitates a repeated query, where the query values may change as a result of interactive adjustments, seeking the appropriate values to convey the message. This type of query benefits from storage of new linear spatial objects, particularly if new linear spatial objects are defined using every point at which the relevant attribute values change.

Intergraph provides two dynamic segmentation products in the MGE system environment. Segment Manager is based on a relational data model, while Dynamo is based on an object-oriented data model.

Both systems employ the same underlying linear data model that relates reference markers to linear spatial objects, and both have:
- interpolation between reference markers, called milemarkers, and
- linkage to attribute data distributed among separate DBMSs using distance referencing schemes.

In Dynamo, a route is an object consisting of a variable-length array of values and milepoint limits for each attribute, and is related by milemarkers to linear spatial objects.

The ARC/INFO implementation of dynamic segmentation relates routes and sections to chains by means of route-attribute and route-section tables. ARC/INFO employs a linear data model wherein routes are related to sections and sections to chains (arcs). Each section has a beginning and ending mile-point value. Also, each section is defined in terms of the percentiles of the arcs on which they lie. This allows sections, and consequently routes, to begin or end within an arc. The route-section table enables ESRI to integrate the concepts of routes and milepoints as an extension of the existing ARC/INFO data model.

The GDS/DSIM application creates new GDS objects for each incident it maps. It does not affect the GDS objects that constitute the chain (route in our nomenclature). Thus routes are collections of GDS objects, which we refer to as sections.

GDS/DSIM supports three concepts for spatial referencing.

---

**TABLE 1.**
A Comparison of Vendor Features and Terminology for Dynamic Segmentation

<table>
<thead>
<tr>
<th>Vendor: System</th>
<th>ESRI:ARC/INFO</th>
<th>Caliper:TransCAD</th>
<th>Intergraph:MGE</th>
<th>EDS:GDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic Segmentation Module</td>
<td>Dynamic Segmentation</td>
<td>Milepost Geocoding/ Dynamic Segmentation</td>
<td>Dynamo Segment Manager</td>
<td>DSIM</td>
</tr>
<tr>
<td><strong>Geographic Data Type</strong></td>
<td><strong>Highway Data Type</strong></td>
<td><strong>Milepost Data Type</strong></td>
<td><strong>Venue Manager</strong></td>
<td><strong>Object Data Type</strong></td>
</tr>
<tr>
<td>Polyline</td>
<td>Section</td>
<td>Section</td>
<td>Section</td>
<td>Section</td>
</tr>
<tr>
<td>Chain</td>
<td>Nodes and Link</td>
<td>Arc</td>
<td>Segment</td>
<td>Chain</td>
</tr>
<tr>
<td>Path</td>
<td>Route or Control</td>
<td>Route</td>
<td>Route/Control Section</td>
<td>Route</td>
</tr>
<tr>
<td>Point Feature</td>
<td>Incidents</td>
<td>Point Event</td>
<td>Milemarker</td>
<td>Offset Distance</td>
</tr>
<tr>
<td>Linear Feature</td>
<td>Segmented lengths, Constant value</td>
<td>Line Event</td>
<td>Line</td>
<td>Line</td>
</tr>
</tbody>
</table>

**System Features**
- New Linear Spatial Objects
  - for pan and zoom (Yes)
  - for spatial analysis (Yes)
- Handles Special Cases
  - proportions meas., error equations (Yes)
  - divided highways (Yes)

* Additional step of line overlay can be used to create new spatial objects for spatial analysis.

104  URISA Journal/Reviews
The first is to use the graphics (sections) as the referencing system. The length of each section is used to determine the x,y of "milepoints" or location of incidents (i.e., 10 miles on the graphics (spatial entity) equals 10 miles relative to the attribute data). The second referencing concept lets the user "define" the length of the route. In this mode the "length" to be dynamically segmented is specified by the data, rather than by its scaled graphical length. The third referencing concept lets the user "define" the length of a section by specifying a starting and an ending value for the section.

One of the problems in comparing approaches is inconsistent terminology. Other problems are incomplete descriptions of approach, different approaches, and different interpretations of the dynamic segmentation problem. Table 1 provides a comparison of approaches and terminology for the above mentioned vendors.

The initial comparison of vendor approaches to dynamic segmentation is incomplete. This is due in part to incomplete available documentation. A more complete comparison would require developing a comprehensive set of evaluation criteria and exercising the systems on comparable problems.

Conclusion

Dynamic segmentation is a solution to the problem of managing linear data in a GIS. It is particularly important for transportation applications. This analysis of dynamic segmentation for transportation applications has identified requirements and approaches. This sets the stage to formulate an evaluation of the different approaches of vendors. The milestone linear data model was developed to help clarify the issues and to provide guidance to GIS developers and users. It may also serve to provide a framework to evaluate various approaches.

This analysis shows that vendors have responded to the demand for dynamic segmentation in different ways. Although each vendor has interpreted the problem differently and has invented slightly different solutions, the differences in data models are not as great as the differences in terminology. This is not an unexpected result, given that the transportation community has not articulated its needs systematically, nor in concert.

An important issue in a systematic and comprehensive solution to the management of highway data in a GIS is whether transportation organizations are willing to adopt a milestone linear data model that impose conditions, such as sections terminating at the end points of chains. More robust solutions are possible, but at a cost. Currently, vendors rather than transportation information users are making these decisions. The transportation user community needs to organize its requirements more systematically. With users and vendors cooperatively defining needs and approaches, standards can emerge that enable differences among software implementation to be evaluated by how well they serve the particular requirements of various dynamic segmentation applications.

Acknowledgments

The authors wish to acknowledge suggestions and encouragement of individuals including: Peter Van Denmark, David Moyer, Timothy Nyerges, and Alan Vonderohe. In addition, vendors have been open and helpful. Specifically, we would like to acknowledge the comments and suggestions of Eric Ziering, Caliper Corporation; Peter Girard, ESRI; James Mohan, EDS; and Joe Francico, David Glenn, and Lynn Williams, Intergraph.

Note

An earlier version of this paper was presented at the GIS-T '92 conference in Portland in Feb. 1992. The data model section of this paper has been expanded subsequent to the GIS-T '92 conference to facilitate comparison of approaches.

The paper is reprinted from the GIS-T '92 Proceedings with permission.

References


Dueker/Vrana/URISAJournal 105
Taking the TIGER by the Tail: A Software and Data Review
Alan A. Lew and Richard E. Klosterman

The introduction of TIGER/Line™ files with the 1990 Census of Population and Housing has presented a new realm of opportunity for socio-economic mapping and GIS applications. The TIGER/Line files provide a comprehensive digital map database of the entire United States tied to 1990 Census data. Both the TIGER/Line files and the accompanying 1990 Census data files are currently available in ASCII format on CD-ROM in most major libraries throughout the country. Demographic data are easily handled with spreadsheet, database and other programs, but TIGER/Line files usually must be converted into a geographic format to be used with mapping software. There are essentially four ways in which this can be accomplished:

1) By having an in-house programmer or consultant write translation routines to meet the specific software and application needs of an organization. This option is not discussed in this review.
2) Through third-party vendors who translate the raw TIGER/Line files into standard formats used by GIS/mapping programs. These include many GIS vendors who provide this service for users of their software. These vendors may even make some improvements to the raw TIGER/Line files. This option is discussed briefly at the end of this review (see Table 3a).
3) Graphic programs that can directly translate the raw TIGER/Line files. Several GIS/mapping programs have this capability (e.g., Caliper’s GISPlus™) or provide it at an extra cost (e.g., Strategic Mapping’s Atlas Import/Export™). Two graphics programs specifically written to work with TIGER/Line files are reviewed here. GeoSight Factfinder™ (Sammamish, Inc.) is a mapping program which reads and translates TIGER/Line file information. Census Windows: Tiger Tools™ (GEOVISION, Inc.) is a display program providing direct access to TIGER/Line files.
4) Third-party translation software. These types of programs typically translate raw TIGER/Line files into Autodesk’s AutoCAD™ .DXF (drawing interchange file) format, as well as selected other formats, which are then read and translated by other GIS/mapping software. This review examines three programs from Micro Map & CAD which translate raw TIGER/Line data into formats that can be interpreted by various graphics software.

There are many choices available for gaining access to TIGER/Line files for use in mapping and GIS (see also the sidebar review). The discussion of these representative products will raise some of the more important considerations in the translation and utilization of TIGER/Line files.

TIGER/Line File Format

The Topologically Integrated Geographic Encoding and Referencing (TIGER) system is an automated geographic data system created by the U.S. Bureau of Census to support the mapping and related geographic activities for its census and survey programs. The TIGER database contains digital data for all 1990 Census map features (e.g., roads, railroads, and rivers), census geographic units (e.g., tracts and blocks), political units (e.g., cities and townships), feature names, and 1980 and 1990 Census geographic area codes for the entire United States. It also contains address ranges and ZIP codes for streets in the urban portion of most metropolitan areas. The core of the TIGER system consists of 3,286 county partition files (some of which cover more than one county); in which geographic coordinates, identification codes, and spatial relationship data are contained.

Under an agreement with the Census Bureau, the U.S. Geological Survey scanned or digitized four categories of carto-
graphic information—roads, hydrography, railroads, and miscellaneous transportation features—for the 48 coterminous states and the District of Columbia from its new 1:100,000 scale map series. Meanwhile, the Census Bureau extended and updated the features in the 1980 GBF/DIME files that contained digital descriptions of the streets and other map features for the urban core of 345 metropolitan areas. The TIGER/Line files currently available are an extraction from the complete TIGER database maintained by the Census Bureau. In the following discussion “TIGER” refers to the original database, while “TIGER/Line” refers to the extraction.

The TIGER/Line extract is a database that is organized as a topologically consistent line network representing the line features traditionally found on a paper map. The extract contains digital data on census geographic areas and associated Federal Information Processing Standards (FIPS) identification codes (e.g., tract and block numbers), municipal and voting district boundaries, and physical features such as roads, streets, and waterways. Each record in the database corresponds to a line segment for a particular map feature (e.g., a street segment from one intersecting street to another or a portion of a river).

The TIGER/Line data includes the latitude/longitude coordinates for the endpoints and shape points of all line segments and the coordinates for all point features. Directions are assigned to each line segment by choosing one of the endpoints as the from node and the other as the to node. Proceeding along the segment in this way allows the system to identify the census geographic areas (e.g., tract and block), political jurisdictions, and so on lying to the left of the segment and areas lying to its right. Address ranges and ZIP codes are also provided for each side of street segments that were included in the 1980 GBF/DIME files; address ranges and ZIP codes are not provided for other areas. Shape records are also included, providing coordinate values that describe the shape of feature segments that are not straight.

The boundary conversion programs considered in this review use the information on the census geographic areas and political jurisdictions lying to the left and right of each line segment to create the desired boundary files. To create the

---

**TIGER Massage from MedGraphics of Chapel Hill NC (919-929-2878)**

TIGER Massage, which was in beta testing when I used it, is a Macintosh program that can build borders for the 1990 and 1980 areas designated in both TIGER/Line Record Types 1 and 3. The TIGER/Line files may be in the precensus or census version, and may be read from CD-ROM as well as from hard disk. The shape points in Record Type 2 can be used as an option.

The output formats are both ASCII text files, The BNA (Boundary ASCII) format, compatible with all Strategic Mapping, Inc. programs, contains a header line with a primary name (to uniquely identify each area), a secondary name (useful for labelling the area), and the number of coordinates in the border, followed by lines with those coordinates. TIGER Massage creates a primary name that is the concatenation of the codes for all of the levels of geography above the area, plus the area’s own number. For an example census tract, “3602315301” would be New York (36), Monroe County (053), Census Tract 153.01; the secondary name would be “15301.”

The second output choice, MDIF (MapInfo Data Interchange Format), is compatible with the versions of MapInfo for graphical user interface available from MapInfo Corp. The format is similar to BNA but has two files, one for the geographic data (.MIF, with ID and coordinates) and one for the attribute data. (.MID, with ID and name). The .MIF files also has a header section that defines the structure of the .MID file, the latter contains only data records, which can have more attributes than the two that TIGER Massage stores.

Peter Van Denark
Software Reviews Editor
boundary file for a particular census tract, the conversion software must first select all of the line segments that have the tract’s identification code on one side and another tract’s code on the other side; line segments that have the same tract IDs on both sides are internal to the tract. The program must then retrieve the coordinates for the “from” and “to” nodes and shape points of all of these segments and chain them together in the correct sequence to define the tract’s boundaries. The program must then recognize when it has finished tracing the tract’s boundaries and store the coordinate points defining the tract’s boundaries in the desired format. It can then repeat the process for all of the remaining tracts in the county’s TIGER/Line file. The process, while conceptually and computationally straightforward, is rather time consuming, explaining why it takes the conversion programs quite a bit of time to generate boundary files for a large number of spatial entities.

The geographic coverage for a TIGER/Line file is a county or statistically equivalent area (i.e., a multicounty area in some states). The TIGER files cover the entire United States, Puerto Rico, the Virgin Islands, and outlying areas such as American Samoa and Guam. The complete set of files contains latitude/longitude coordinates for more than 25 million feature intersections and end points that define nearly 40 million feature segments; the total file size is approximately 25,000 Mb. The average file size for a state or statistically equivalent area in 500 Mb, although they range from fewer than 4 Mb to approximately 1,800 Mb. The average file size for a county or statistically equivalent area is 7.5 megabytes, with a range from less than 1 Mb for Jemo Island in the Marshall Islands to approximately 125 megabytes for Los Angeles County, California.

Although conceptually a single file, the TIGER/Line records for a county actually consists of 12 “record types” that collectively contain all of the geographic information for the county. The 12 record types are stored as separate files with contents that can be combined as necessary to serve different purposes. The contents of each record type are described thoroughly in the TIGER technical documentation (see table 3b) and elsewhere (e.g., Van Dam 1992; Klosterman 1992). Only the first three Record Types are used by the software considered in this review.

Record Type 1, the Basic Data Record, provides a single record for each unique line segment in the Census TIGER/Line file. Each record contains 41 data elements including:

1) A unique record number identifying the line segment,
2) The name of the feature,
3) The latitude/longitude coordinates of the segment’s “from” and “to” nodes,
4) FIPS code identifiers for the political subdivisions on the left and right of the segment,
5) 1990 Census geographic area codes for each side of the line segment, and
6) Address ranges and ZIP codes for each side—in metropolitan areas where these are available.

Record Type 2, Shape Coordinate Points, contains the latitude and longitude coordinates for up to 10 additional “shape” points that describe the shape of each line segment in Record Type 1 that is not straight. Record Type 3, additional Censu Geographic Area Codes, contains the 1990 voting district codes (state, county, MCD/CCD, and place) and additional 1980 and 1990 Census geographic area codes for the left and right side of each line segment identified in Record Type 1.

It is important to recognize that the information provided in the TIGER files is designed for data collection and analysis purposes and provides only a rough approximation to actual locations on the land. The positional accuracy of the points and lines in the TIGER files is no greater than the accuracy standards of the 1:100,000-scale USGS maps (i.e., at least 90 percent of points are within 164.04 feet of their actual location). Data recorded at this scale are not suitable for high-precision measurement applications such as property transfers and engineering or construction projects. However, the network and census geographic area information provided by the TIGER/Line files can be used with census or local attribute data and generally available GIS software to perform a wide range of spatial analysis activities of particular interest to planners and other public officials.

Potential TIGER/Line file applications include: thematic mapping, address matching,
network analysis, and spatial analysis. For example, the TIGER/Line files can be used in a GIS or desktop mapping program to display block boundaries along with locally defined districts (e.g., planning districts or traffic analysis zones). Standard GIS overlay operations can then be used to determine the portion of each block's area that corresponds with the population, housing units, etc., that lie within each planning district. This allows census information that was initially available only for census geographic boundaries to be analyzed for locally defined areas.

**Translation Software Overview**

Micro Map & CAD provides both converted TIGER/Line file drawings (Table 3a) and software for end-users to convert their own TIGER/Line files (Tables 1 and 2). Three software conversion programs are available:

- TigerDXF™ translates TIGER/Line files into Drawing Interchange File (.DXF) format for use by most computer-aided design (CAD) programs. Each TIGER feature class is placed on a layer. Additional layers further separate features according to their source (GBF/DIME, DLG, etc.). Layer colors and line types may be defined for the conversion process. Line segments are converted into polylines for which line widths can also be predefined. (A similar TigerMML translation program for MapInfo™ is also available.)
- TgrBndry™ extracts political and census statistical area boundaries. Each area's boundary comprises a single polygon and each polygon is on a separate layer named after its FIPS identification code. Three conversion formats are supported: .DXF (for CAD programs), .BNA (for ATLAS Import/Export, as well as ESRI's ArcInfo™), and .MBI (for MapInfo). Micro Map & CAD's TgrBndry also checks for polygon closure. Non-closures and other boundary anomalies are written to a separate log report file.
- The three Micro Map & CAD products are designed for graphic conversion only. Map display and editing capabilities are provided by the CAD or GIS application program which will import the converted TIGER/Line file. These products were the only conversion programs reviewed which provided a visually more accurate UTM projection, in addition to TIGER's non-projection latitude and longitude. In addition, TigerDXF and TgrNames both come with a LISP program which enables direct access to the original raw TIGER/Line data from within AutoCAD. This provides easy access to the non-graphic attribute information in TIGER.

The strength of these programs is in the depth of detail provided in the converted files. Unlike the other products reviewed here, each feature class from each source base has its own layer. Importing these into powerful CAD and GIS programs allows for maximum flexibility and application development. The downside of all these layers is the difficulty of managing them. The Census Feature Class Code is unnecessarily detailed for many applications.

**GEOVISION's Census Windows: TIGER TOOLS** was one of the first programs designed, in part, for libraries and other repositories of Census TIGER/Line CD-ROMs. Like the Micro Map & CAD products, TIGER TOOLS is essentially a processing utility for TIGER/Line files. However, TIGER TOOLS also includes facilities for viewing, editing, and linking TIGER/Line files to a variety of other programs. Much of its strength comes through its operation in the Microsoft Windows™ environment. Windows allows TIGER TOOLS to link TIGER/Line files to a variety of other software applications. TIGER TOOLS creates device-independent bitmapped images (.BMP) from raw TIGER/Line data, which may be copied to most Windows and DOS paint programs. With Window's Dynamic Data Exchange (DDE) capabilities, TIGER TOOLS is able to connect the TIGER/Line database to Windows spreadsheets (especially Microsoft Excel™), database programs, word processors, and graphics programs. TIGER TOOLS, however, is limited to exporting the TIGER/Line map attributes to other programs, it cannot import data to create, for example, thematic maps.

Bitmap map images created from TIGER/Line files are
<table>
<thead>
<tr>
<th>Vendor</th>
<th>Geovision</th>
<th>Sammamish</th>
<th>Micro Map &amp; CAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td>TIGER TOOLS v. 3.03</td>
<td>FactFinder v.4.1c</td>
<td>TigerDXF v.1.2</td>
</tr>
<tr>
<td>Version Reviewed</td>
<td>MS-DOS/Windows</td>
<td>MS-DOS</td>
<td>MS-DOS</td>
</tr>
<tr>
<td>Operating System</td>
<td></td>
<td></td>
<td>TgrNames v.1.2</td>
</tr>
<tr>
<td>TIGER/Line Versions Read</td>
<td></td>
<td></td>
<td>TgrBdary v.2.52</td>
</tr>
<tr>
<td>.F1? Prototype</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>.F2? Pre-Census</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>.F3? Initial Voting Districts</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>.F4? Post-Census</td>
<td></td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Area Extraction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Census and Political Boundaries</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Selected Place Code Boundaries</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Latitude/Longitude Rectangle</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Multiple County</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Multiple Boundary Levels</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Entity Extraction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political and Census Area Polygons</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Single Feature Class Code</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Multiple Feature Class Code Groups</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Names on Drawing</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Record Type 2 (shape) optional</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Projection of Extraction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latitude/Longitude</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>UTM</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Multiple or Batch Translation</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>in single command</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Display of Map</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zooms and Pans</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Separates entities on layers or by colors</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Saves Display as Bmp</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Saves and recalls displayed maps</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Display of Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Census Demographic Data</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>User Imported Data</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Editing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw TIGER/Line Files</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Drawing Features</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Demographic Attribute Data</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Printing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Map</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Raw TIGER/Line Data</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>HPGL Plotter Support</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Save Drawing Formats</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.DFX - AutoCAD</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>.BNA - Atlas&quot;GIS&quot;</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>.MBI - MapInfo</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Save Raw Data Formats</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASCII Delimited file</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>.DBF</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>.XLS (Microsoft Excel)</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>

1Requires transformation of raw data into delimited file format
2Requires query of individual streets after extraction
3All Micro Map & CAD products rely on other software to display, edit, and print
4Can be accomplished with separate conversion.
5Maximum of 12 layers
6Uses Windows Print Manager printer drivers
7Requires optional Atlas Import/Export module

110 URISA Journal/Reviews
TABLE 1b.
Comparison of TIGER/Line File Access and Translation Software

<table>
<thead>
<tr>
<th>OTHER LINKAGES AND CAPABILITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TIGER TOOLS:</strong></td>
</tr>
<tr>
<td>Raw TIGER/Line data to Microsoft Windows programs via Windows Dynamic Data Exchange (DDE) linkage, with special link to Microsoft Excel spreadsheet. Also saves in .GEO format for use with GEOVISION’s StatMap III.</td>
</tr>
<tr>
<td><strong>FACT FINDER</strong></td>
</tr>
<tr>
<td>Can import ASCII fixed or delimited format files, dBase III/IV DBF files and Lotus.WK7 files as attribute data to link with the map drawing. FactFinder can also: calculate map distance; create boolean thematic maps based on attributes; add and delete lines, text, symbols, fill areas, and new features on a map; and generate reports.</td>
</tr>
<tr>
<td><strong>TIGERDXF and TGRNAMES:</strong></td>
</tr>
<tr>
<td>AutoCAD AutoLISP program to query raw attribute data and insert as text in the AutoCAD drawing using TIGER record numbers as AutoCAD entity handles. This requires direct access to the raw TIGER/Line files.</td>
</tr>
</tbody>
</table>

**DOCUMENTATION**

| **TIGER TOOLS:** |
| Adequate, some errors (recent changes not included in the manual); No lists of error messages; Very good TIGER/Line file description and reference |
| **FACT FINDER** |
| Very clear and comprehensive; Each module has different manual; Basic TIGER/Line reference, but does include all FIPS State and County codes |
| **TIGERDXF, TGRNAMES, and TGRBDXRY:** |
| All three: Minimal, but adequate; Some steps missing; No list of error messages; Basic TIGER/Line reference |

TABLE 2.
TIGER/Line File Software Products.

Software reviewed are indicated by (*)

<table>
<thead>
<tr>
<th>Software reviewed are indicated by (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Micro Map &amp; CAD</strong></td>
</tr>
<tr>
<td>*TigerDXF Translator</td>
</tr>
<tr>
<td>*Tgrbdry - TIGER Boundary Extraction</td>
</tr>
<tr>
<td>*Tgr:NAMES - TIGER Line &amp; Names</td>
</tr>
<tr>
<td>TigerMMI Translator</td>
</tr>
</tbody>
</table>

GEOVISION, Inc.

<table>
<thead>
<tr>
<th>Software reviewed are indicated by (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Census Windows: TIGER TOOLS</td>
</tr>
<tr>
<td>TIGER CAD Translator</td>
</tr>
<tr>
<td>StatMap III for Windows</td>
</tr>
</tbody>
</table>

Sammamish Data Systems

<table>
<thead>
<tr>
<th>Software reviewed are indicated by (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>*GeoSight FactFinder (U.S.)</td>
</tr>
<tr>
<td>GeoSight FactFinder (one state)</td>
</tr>
<tr>
<td>*TIGER/LINE Extraction module</td>
</tr>
<tr>
<td>Public Law 94-171 Extraction module</td>
</tr>
<tr>
<td>Summary Tape File 1A Extraction module</td>
</tr>
<tr>
<td>Summary Tape File 3 Extraction module</td>
</tr>
<tr>
<td>Geographic/Map Files Extraction module</td>
</tr>
<tr>
<td>GeoSight</td>
</tr>
<tr>
<td>GeoSight Professional</td>
</tr>
</tbody>
</table>
TABLE 3a.
Selected TIGER/Line Map Suppliers.

<table>
<thead>
<tr>
<th>Supplier</th>
<th>7.5&quot; Quadrangle Maps</th>
<th>5.5&quot; Quadrangle Maps</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Digital Cartography (ADC)</td>
<td>- Available in UTM (meters) or State Plane Coordinate System (feet)</td>
<td>- Available in Latitude-Longitude and UTM projections</td>
</tr>
<tr>
<td></td>
<td>- Formats: .DWG and .DXF (both AutoCad), or DGN (Intergraph)</td>
<td>- Formats: .DXF (AutoCad), .BNA (Atlas*GIS), or .MBI (MapInfo)</td>
</tr>
<tr>
<td></td>
<td>- Contains all TIGER feature classes separated by layer</td>
<td>- Pricing depends on size of file</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>County Maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$150 per Quad + $25 to $75</td>
<td>$30 to $100 per Quad</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micro Map &amp; CAD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sammamish Data Systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>County Maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$150 to $300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$200 to $400</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$300 to $800</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$300 to $800</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$200 to $700</td>
<td></td>
</tr>
<tr>
<td>Also available: DXF translators and translated AutoCAD drawings from: DLG, DEM, WDB, Etak, GDT, and MOSS; Miscellaneous World and U.S. AutoCAD drawings; SPCS coordinate translator, ZIP code maps</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

single layers of a requested extraction, although several extractions can be overlain in different colors. Each extraction results in the creation of a completely new set of TIGER/Line files containing the same data structure as the original, but with only the extracted entities. With TIGER TOOLS, it is possible to create a TIGER/Line file for a single municipality, by extracting all of the feature codes contained within a particular place code. TgrBndry can also do this, but only for polygon boundaries, not for line segments. The other program reviewed here do not have this capability. TIGER TOOLS is also the only one of the reviewed programs that pro-
vides the ability to edit the raw TIGER/Line data and exporting them to .DBF (dBASE™) format. In addition, TIGER TOOLS is able to read earlier versions of the TIGER/Line files, although it must first convert them into a delimited format in which each record is separated by a carriage return (post-census TIGER/Line files are already in this format).

The ability to export TIGER/Line extractions into .DXF files is provided, but is not considered a core function. .DXF output contains a maximum of eight layers, each assigned to a major TIGER feature class code. By manipulating the initial extractions, however, it is possible to create many more layers in a selective manner. GEOVISION also offers a non-Windows TIGER/Line file converter, known as TIGER CAD Translator™. Its features are the same as the .DXF conversion capability of TIGER TOOLS.

All of the translation programs convert each line segment into a separate polyline. A feature unique to TIGER TOOLS is the ability to either keep line segments as separate polylines, or to link all line segments with the same name together as a single polyline. The new polylines become added features to the raw TIGER/Line file database. Like TgrBndry, TIGER TOOLS check for missing segments when building polygons. However, if found it will automatically fix the boundary error without notifying the user.

GeoSight FactFinder from Sammamish Data Systems is the newest TIGER/Line file product on the market. It is a scaled down version of the GeoSight™ GIS program and includes a module specifically designed for extracting, displaying, and manipulating TIGER/Line files. The U.S. version of FactFinder comes with predefined state, county, and city maps. State-level FactFinder is available with county, city, and census tract maps for individual states. Both versions include the PL94–171 basic 1990 Census data, as well as selected data from the City and County Data Book, 1988 (U.S. Bureau of the Census, S/N 003–024–06709–9). The TIGER/Line extraction module enables users to expand upon the maps that come with the version of FactFinder purchased.

FactFinder displays TIGER/Line maps and links them to census demographic data. Users can also input their own attribute data and tie them to the maps. Additional extension modules are available for other types of 1990 Census mapping and demographic data (see Table 3a). As with TIGER TOOLS, FactFinder’s .DXF conversion capability is ancillary to its functionality. FactFinder can also import .DXF files.

FactFinder comes with most of the tools one would expect in a basic GIS, including map editing, thematic mapping, report writing, layering, and boolean and mathematical attribute manipulation. Its major limitation is that it allows only 12 layers, resulting in TIGER feature classes being aggregated (e.g., all roads on one layer). While this is an improvement over TIGER TOOLS, the Micro Map & CAD products may provide much more functionality for more versatile applications. The full GeoSight GIS does, however, permit an unlimited number of layers.

Tiger Translation Considerations

Time and space: For all of the TIGER/Line file translation programs, time and disk space are crucial operating environment factors. The TIGER/Line file for a large county, such as Coconino County, Arizona (66Mb) requires several hours (up to 12 on a 386DX running at 25MHz) to work through the calculations to convert the entire database into a .DXF file. Extractions of smaller, sub-county areas or only specific features required less time. FactFinder and TIGER TOOLS operate at approximately the same speed. Both are much faster than the Micro Map & CAD products. A nice feature of FactFinder is its inclusion of a “percent finished” indicator as well as shaded bars to show how much progress has been made in the conversion.

The resulting .DXF file for all feature classes for Coconino County created by TigerDXF was approximately 44Mb. Since even small counties can be over 5Mb in size, some operations will require considerable computer running time. For larger counties disk space becomes an important consideration. To help alleviate these problems, the Census Bureau will soon release the TIGER/Boundary™ extraction files (see Table 3b). These files will be in two forms, including a “thinned” version with fewer coordinate points.
TABLE 3b.
U.S. Bureau of the Census TIGER Related Products

<table>
<thead>
<tr>
<th>CD-ROMs:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-P.L. 94-171 (basic 1990 Census</td>
<td>$150</td>
<td>-for one CD-ROM</td>
</tr>
<tr>
<td>data)</td>
<td>$1000</td>
<td>-complete set of 10 CD-ROMs for entire U.S.</td>
</tr>
<tr>
<td>-1990 Census TIGER/Line Files</td>
<td>$250</td>
<td>-for one CD-ROM containing several states</td>
</tr>
<tr>
<td>-1990 Pre-Census TIGER/Line Files</td>
<td>$100</td>
<td>-each</td>
</tr>
<tr>
<td>-Housing, Agriculture, &amp; Economic</td>
<td>$150</td>
<td>-more comprehensive Census demographic data than P.L. 94-171</td>
</tr>
<tr>
<td>-Summary Tape File extractions</td>
<td>$150</td>
<td></td>
</tr>
</tbody>
</table>

Most Census CD-ROMS are available at major university libraries and state data centers.

<table>
<thead>
<tr>
<th>Magnetic Media:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-TIGER/Geographic Reference File-</td>
<td>$275</td>
<td>-names and geographic codes (to be replaced by TIGER/GIS in 1992)</td>
</tr>
<tr>
<td>Names</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-TIGER/Census Tract Street Index</td>
<td>$25 to</td>
<td>-index to match street address to census tracts</td>
</tr>
<tr>
<td>-Tiger/Database</td>
<td>$175</td>
<td></td>
</tr>
<tr>
<td>-TIGER/SDT</td>
<td>1992 release</td>
<td>-expanded version of TIGER/Line Files from the complete</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-TIGER/Files in FIPS Spatial Data Transfer Standard format</td>
</tr>
<tr>
<td>-TIGER/Boundary</td>
<td>1992 release</td>
<td>-polygon boundary data only</td>
</tr>
</tbody>
</table>

TIGER Publications:
TIGER: The Coast-to-Coast Digital Map Data Base (1990)

TIGER Help Lines at the Bureau of the Census:
Applications and Training: State and Regional Programs
Future TIGER Products: Geography Division (301) 763-1580
Geographic Concepts: Geography Division (301) 763-4664
Computer Mapping: Geography Division (301) 763-5720

designed for desktop mapping applications.

**TIGER/Line versions:** All of the programs work with the post-census TIGER/Line files (Table 1). The version of the files is indicated by the middle location in the filename extension, with .F4? being the final 1990 Census version. There are up to 12 different record types for each county; however, the translation programs typically only require record types .F41 (basic), .F42 (shape points), and sometimes .F43 (additional codes). These three record types comprise from one-third to two-thirds of the size of the entire TIGER/Line file set for a county. Some of the translation programs can process earlier TIGER/Line file versions, as well. The need to work with earlier TIGER/Line files, however, is likely to be limited for most applications.

**Extraction:** TIGER/Line files are packaged by county. All of the programs reviewed here operate at the basic county level. Only TIGER TOOLS and Tgr-Bndry (for polygon boundaries only) provide the ability to select specific FIPS Place Codes to narrow the operating level to a political or census unit below the county. Instead of Place Codes the other products allowed a latitude/longitude rectangular area to be defined at the sub-county level. While this is adequate for most purposes, it would be less effective for an awkwardly shaped district. TIGER TOOLS and FactFinder both allow multiple boundary types to be selected in a single extraction (e.g., both census tracts and blocks). Census places which cross county boundaries can also be problematic. In addition to selecting a portion of one county, FactFinder allows extraction of multiple counties or a latitude/lon-
gtitude rectangle that overlaps more than one county.

Each line in a TIGER/Line file potentially serves two purposes: as a linear feature (road, river, etc.) and as part of a polygon border (voting district, census tract, city, etc.) A separate conversion of the TIGER/Line files is required for each type. Extracting all of the feature classes included in the TIGER/Line database can create very large, and potentially unmanageable, files. Most of the conversion programs allow individual or multiple features to be extracted, making it easier to manage the process of creating layers. For example, FactFinder groups the 179 feature class codes into nine types for selection purposes: Freeways, Highways, Arterials, Street/roads, Railroads, Shorelines, Streams, Other visible, and Non-visible. Users with database may, however, modify the content of these groupings.

TgrNames actually places the names of streets on the map. While this is a useful mapping utility, it is not necessarily desirable for GIS. FactFinder accesses the street names in a more typical GIS approach by doing a query to the raw TIGER/Line database. Through this query, the names can be edited and added to the map. In order to reduce the size of drawing files, TIGER TOOLS allows conversions to exclude the Record Type 2 shape points. None of the programs reviewed here contain other mechanisms for file size reduction. This, in particular, is an area in need of further development.

**Projections, displays, printing, and editing:** The three Micro Map & CAD programs offer no display capability, but do offer greater projection control and superior .DXF file formats for use with CAD programs. FactFinder offers an impressive array of display, edit, and printing features, while the strength of TIGER TOOLS is in its linkages to other Microsoft Windows applications. TIGER TOOLS allows editing of the raw TIGER/Line database, while FactFinder is the only program reviewed here which can import demographic and other user-defined attribute data to create thematic maps.

**Obtaining Tiger/Line Files and Census Demographic Data**

A variety of vendors provide TIGER/Line files in formats that may be directly read or translated by most of the GIS and mapping programs currently in use. Strategic Mapping, Inc. sells street files for its Atlas GIS™ and Atlas Pro™ programs and MapInfo Corp. provides similar products for its MapInfo mapping software. Micro Map & CAD, Sammamish Data Systems, and American Digital Cartography all provide converted TIGER/Line file data in .DXF and other file formats (Table 3a). It is also possible to buy TIGER/Line files directly from the U.S. Bureau of the Census or to obtain them from major libraries and data centers.

TIGER/Line files are distributed by the U.S. Census Bureau on CD-ROM. These files are in the public domain and may be redistributed freely once a copy is obtained. All of the programs reviewed here access and convert TIGER/Line files both on CD-ROM and on a computer's hard disk. Most major public
and university libraries and state data centers have the Census CD-ROMs. Not all of these agencies, however, provide a convenient or inexpensive means for accessing county TIGER/Line files.

The process of copying the ASCII TIGER/Line files from CD-ROM is simple enough. All of the files for each county are stored in a MS-DOS subdirectory under a directory for each state. The directories are named using the FIPS codes, which generally number states in two-digit alphabetical order (e.g., Arizona = 04), and counties in three-digit alphabetical order (e.g., Coconino County, AZ = 005). The problem arises in moving the county TIGER/Line files from the host computer to the user’s computer. A tape drive is not always available, and floppy drives have insufficient disk space to hold the larger files.

One way to overcome this problem is to use the MS-DOS utilities BACKUP.EXE and RESTORE.EXE to split large files within a subdirectory among several disks. However, there are some potential problems with BACKUP and RESTORE in that they are dependent on particular versions of MS-DOS and earlier versions of them have a reputation of being unreliable.

Alternatively, there are inexpensive shareware programs which can also help. The first is the file compression program PKZIP (PKWARE, Inc.). Because of the repetitive nature of the TIGER/Line data, PKZIP is able to compress them to approximately one-sixth their original size. Even this, however, may still be too big for a 1.44Mb floppy disk. A free program called CHOPPER™ (Dlugosz Software) alleviates this problem by breaking the large compressed file on the hard disk into smaller pieces which can then be copied to a floppy disk and reassembled on another computer.

Purchasing pre-processed TIGER/Line files is another way of obtaining access to TIGER (Table 3a). In addition to their three conversion programs, Micro Map & CAD sells converted 7.5" quadrangle maps and county-wide maps with all of the features outlined in Table 1. American Digital Cartography (ADC) provides the same basic TIGER/Line file conversions discussed above at the 7.5" quadrangle level. In addition ADC offers enhancements such as double-lined roads, elevation contours, Public Land Survey Section lines, geodetic control points, and conversion to the 1983 North American Datum. Both of these vendors also provide a wide range of other maps for GIS and planning applications.

A variety of other TIGER extraction files are also being developed in CD-ROM and tape formats by the Bureau of the Census (Table 3b). These include the TIGER/Database™ files, which are an expanded version of TIGER/Line, and the TIGER/Boundary files, which is a smaller version for desktop thematic mapping purposes. While TIGER/Line files have the potential to be used for a wide range of applications in and of themselves, they are particularly well suited to be used with the 1990 Census demographic data. The PL94–171 CD-ROM contains the initial 1990 Census data and provided the basis for political redistricting. The first of the more extensive Summary Tape Files (STF) CD-ROMs are now becoming available, providing a wealth of detailed information that can be easily linked to the census TIGER/Line files.

As discussed above, Sammamish offers additional modules for FactFinder to access and integrate the demographic files with the TIGER/Line files. A separate module must be purchased for each STF (Table 2). GEOVISION recently came out with a new program, StatMap III™, which uses TIGER TOOLS maps to display PL94–171 and STF data, as well as user defined data. Like TIGER TOOLS, StatMap III is a Windows program with two-way DDE linking capacity to other Windows programs. Users of other mapping and GIS programs can tie the census demographic data into their TIGER/Line maps by using TIGER place codes for polygon linkages, while individual features can be linked to external databases via TIGER/Line record numbers.

Conclusions

Since their inception, TIGER/Line files have been viewed as providing a major breakthrough in GIS and computer mapping applications. The Census Bureau recognizes that the TIGER database and related products are extremely useful for applications that go far beyond its own data collection and mapping needs. It
also recognizes that the current TIGER information inevitably contains many errors and is increasingly out of date. As a result, it has undertaken an ambitious effort to continue to update and improve its internal database and to release new and improved data products to the public.

The array of services developed to take advantage of this opportunity is only now beginning to fully develop. The examples presented here, while a first step in that direction, give a good indication of the range of options that are developing. Each serves a market niche. In the final analysis it is up to the end user to determine just how he or she might best take hold of the TIGER’s tail.

Alan A. Lew is associate professor in the Department of Geography and Public Planning at Northern Arizona University. He teaches courses on computer applications in planning to students in the geographic information management degree program. His research interests focus on tourism development, including the application of GIS to tourism studies.


Note

All trademarks are the property of their respective owners.

References


Atlas Pro for the Macintosh

In Depth:

William Casey

Atlas Pro™ for Macintosh® computers is the latest and most comprehensive of a series of Mac mapping programs dating back several years. It was originally published as MapMaker in 1988 by Select Microsystems in Yorktown Heights, NY. Its creative genius, Daryl Scott, and his organization were absorbed by Strategic Mapping, Inc. (SMI) of San Jose, CA in 1989. SMI has been widely acknowledged as a leader in desktop mapping with its Atlas Software product lines for both the IBM-PC compatible and Macintosh markets. SMI also sells many specialized boundary and data files, beyond what it supplies as a standard part of its mapping products.

SMI has since issued two major versions of the Mac product: Atlas*MapMaker version 4.0/4.5 and most recently, Atlas Pro. SMI's view of the market for desktop mapping products involves three distinct tiers. At the bottom are what it describes as purely "presentation mapping" products. These include Atlas MapMaker 4.5 on the Mac (priced at $395; it was $495 before Atlas Pro was issued) and its counterpart for the Windows environment. At the high end is SMI's $2,500 Atlas GIS product, which is available for IBM-PC and compatibles. Atlas GIS includes many major geographic information system functions, and SMI's plans include its availability on the Macintosh side. Atlas Pro retails for $795 and occupies the middle level of the three-tiered model. It has the power of the original MapMaker program with, as we discuss below, additional facility to manage and manipulate associated data and to analyze that data directly from the computer screen's map display.

Atlas Pro for the Mac is surely derived from the original MapMaker program. It might be viewed as a higher and more capable evolution of the original system, or it can be viewed as a wholly new product. Either way, it is a substantial contribution to the inventory of significant Mac-based mapping products.

Desktop Mapping Functionality

A desktop mapping facility, at its most fundamental level, enables display of data on maps. The user chooses the spatial extent of the map (a city, a county, the 50 states) to meet his or her need, and then selects the kinds of data items (attributes) that are to be displayed. The mapping software combines these two fundamental components and produces map displays that seek to communicate effectively with map readers in graphic terms. Display mechanisms such as shaded areas, dots, and point symbols of different sorts are typical means by which this class of programs accomplishes its purpose of map production.

Once desktop mapping software has combined data and spatial descriptions, it can then potentially provide additional analytical power. This is where thematic mapping programs can take on GIS characteristics: when users can isolate (select) data on a map and the system can display and/or analyze data associated with the area.

Atlas Pro Overview

Atlas Pro is an industrial-strength mapping facility, packed with numerous generalized and powerful features that make it attractive to a range of market segments: technical, academic/research, as well as commercial users. In experienced hands, it can produce extremely high-quality thematic maps or simple boundary maps at virtually any scale. The user retains a wide range of control over output characteristics for important map components: boundary lines, fill patterns, point symbols, type, legend op-
Atlas Pro: Summary Table

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Rating</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall thematic features</td>
<td>Excellent</td>
<td>Wide range of function</td>
</tr>
<tr>
<td>Additional Features</td>
<td>Good</td>
<td>Desktop mapping &quot;plus&quot;</td>
</tr>
<tr>
<td>Quality of map output</td>
<td>Very Good</td>
<td>Improved legends would improve rating</td>
</tr>
<tr>
<td>Performance</td>
<td>Excellent</td>
<td>Too many non-standard aspects</td>
</tr>
<tr>
<td>As a Mac application</td>
<td>Satisfactory</td>
<td>Needs professional rewriting and reorganization</td>
</tr>
<tr>
<td>Documentation</td>
<td>Unacceptable</td>
<td></td>
</tr>
<tr>
<td>Ease of learning</td>
<td>Satisfactory</td>
<td>There are many special &quot;tricks&quot; to know, and they take a while to sink in</td>
</tr>
<tr>
<td>Ease of use</td>
<td>Satisfactory</td>
<td>Confusing organization of menus and improving with time.</td>
</tr>
<tr>
<td>Error Handling</td>
<td>Satisfactory</td>
<td>System still freezes on occasion</td>
</tr>
<tr>
<td>Support</td>
<td>Good</td>
<td>Calls generally returned promptly and information missing from manuals patiently explained</td>
</tr>
<tr>
<td>Value</td>
<td>Very Good</td>
<td>For the serious maker of maps</td>
</tr>
</tbody>
</table>

...tions, and so on. The outputs available through Atlas Pro can without doubt be of publishable quality when prepared with care and printed on appropriate output devices (laser printers or better).

Atlas Pro offers users more power than simple thematic mapping capabilities and to that extent deserves being classified as a "desktop GIS," by Peter Van Demark's standards. (See "A Comparison of Four MS-DOS Desktop Geographic Information Systems: . . . ," URISA Journal, Vol. 3, No. 2.) The product "has more functionality than a thematic mapping program" and "allows you to point to a geographic feature on a map display and be shown a data record for it." On the other hand, Strategic Mapping neither advertises or otherwise claims that Atlas Pro is a geographic information system (GIS). Though the package includes some spatial operators (e.g., it permits extraction data from user's screen selections), these are modest in scope. The data side of the Atlas Pro environment consists of several spreadsheet-like tables but does not offer more than that in terms of structuring or handling capacity. Although its layered architecture is attractive for organizing different kinds of boundaries and data, it does not offer comparable advantages on the output side of map production.

Atlas Pro is not without its faults. It remains surprisingly daunting—almost confusing—to use; I state this after working with its predecessor for several years. There are several major functional limitations, its documentation is weak, and there are numerous small, quirky problem areas that one would like to have seen addressed and eliminated over the years. Nonetheless, this complex program is more stable, more powerful, and more customizable than its predecessors and deserves its reputation as one of the premier products of its type on the Macintosh platform.

Atlas Pro is feature-rich—full of all kinds of whistles and bells that make it ideal for a variety of map-related tasks. We will discuss the system in general and cover many of its major features. Be aware, however, that the product contains—generally to its credit—many other facets.

The version of the system we used in preparing this piece began as Atlas Pro version 1.0, and was subsequently replaced by version 1.03. That version was used for the bulk of the review, although it was replaced by version 1.06 as I neared completion. All three versions arrived within a total time span of roughly 75 days. These successive system releases can be viewed in two ways: either early editions of Atlas Pro could have (and, presumably should have) had fewer bugs and other evidences of instability or they reflect a vendor willing to make changes promptly to address software problems.

Sample Atlas Pro Maps

Figures 1 through 4 illustrate four different kinds of maps produced by Atlas Pro. Each illustrates a different kind of thematic capability in the product. Figures 1 and 2 display point data with fixed size and graduated symbols; Figures 3 and 4 involve areal data depicted by shading and dot density. Atlas Pro provides other mapping options, but these four are the most practical and would typically represent the most widely used features of the program.
Overview

There are two fundamental components in the Atlas Pro mapping environment:

- **Geography Files.** These delineate areas or other spatial features that are of interest for a specific mapping problem. Traditionally they have been known as "boundary files"; "geography files" is a somewhat broader term encompassing not just areal boundaries, but also lines (roads, rivers, railroads) and point locations. The scale can be as small as a projection of the earth's continents or as large as needed—say a couple of city blocks.

- **Data Files.** This can be thought of as a spreadsheet-like (or relational database) table. One column contains the key by which Atlas Pro will connect areas and points to the data at hand. Examples of key fields would be state codes for states (AZ, MN, MI), ZIP Codes, or FIPS codes for county level or minor civil division data. Other columns represent variables for display (e.g., population, population density, income). There would be as many rows, potentially, in this table as there are entities to be mapped.

Data can be entered directly into Atlas Pro, but for other than simple applications the user normally wants to bring in data from an outside source. Delimited text files can be read directly, SQL access is provided, and "dbf" format is supported as well. Surprisingly, Excel or other spreadsheet file formats cannot be read directly by Atlas Pro: they must first be written as text-format files and then can be brought into the mapping system using an "Import" command. For some kinds of applications, publish and subscribe can obviate this process, but the ability to read spreadsheet data directly would be a plus.

A "map" in the Atlas Pro environment is not a single printed or plotted output. Instead, a map as it is constructed by the user is an integrated collection of geography files and associated data tables and display formats. From one Atlas Pro "map"—stored and managed as a unit by the system—the user can and usually does generate many differing thematic outputs.

**Geography Files**

Much of Atlas Pro's admirable power lies in its flexibility to map exactly the area needed by the user. Differing boundaries can be combined, even after the map has been constructed. For example, they can be overlaid on each other at map generation time such that county, state and highway outlines can be displayed simultaneously, each with its own line type and weight. These boundaries are selected by the user and organized into layers for convenient management.

Atlas Pro is shipped with a comprehensive collection of such boundaries—countries, U.S. states and Canadian provinces, all U.S. counties, 3-digit ZIP Code regions, major cities, etc.—but the long-term value lies in its capability to map any area of interest. If the user can come up with a boundary file, then he/she can use it in one of several ways and build a customized map for whatever the specific purpose.

Boundary data files can be both imported and exported. An existing boundary that cannot be combined with another due to differing projection methods can easily be "deprojected" back to a text file consisting of named areas, each of which is made up of a series of latitude/longitude points or x/y coordinates. That text file can then be imported and projected within Atlas Pro, resulting in a new boundary file. This facility is the foundation of the program's boundary expandability, since new points, lines, and area data sets can be brought in and projected by the system as needed. Atlas Pro also supports digitizing and creation of boundaries "automatically" from PICT files as additional sources of boundary information.

Several different projections are supported including Miller, Mercator, and other cylindrical projections, and Albers and Lambert's conic projections, each of which permits the user to change standard parallel settings, central meridians, etc. Users of x/y coordinates have their own group of options, including scaling so that Atlas Pro can handle large numbers such as State Plane or UTM coordinates.

Many current commercial marketing efforts are focused on ZIP Codes. Since much business data (as well as increasing amounts of U.S. Census data) are reported at the ZIP Code level, Atlas Pro provides additional capabilities to allow users to view data by ZIP and ZIP Code aggregates. Data coded by ZIP can be translated—reaggre-
gated—by the program to other geographic bases such as Metropolitan Statistical Areas (MSAs), Designated Market Areas (DMAs), Areas of Dominant Influence (ADIs), etc. In addition, user files based on ZIP Codes can be translated and projected into point-based boundary-type files, with Atlas Pro providing

the centroid of each ZIP Code area. These points can then be used as the foundation for maps showing sales or populations by ZIP Code using graduated circles or varying point symbols.

Data Considerations

Data can be associated with each defined layer in an Atlas Pro map. For example, a layer containing county boundaries for Ohio could have one set of attributes associated with it:

Each layer, in turn can have tied to it a data table of multiple columns (attributes or fields), including a designated primary key that relates the data to the geographical unit. In the case of Ohio counties, these might include the three-digit FIPS county code (key), plus—for example—county areas, populations, average incomes, and so on. Only one row of spreadsheet data can be tied to a specific unique place (point or area) key value; duplicates are simply ignored when data are imported. This requires the user to aggregate his/her data at the proper level (say, by county if the mapping is to be done at that level) in advance of introducing it into Atlas Pro.

There is a capacity to join attribute files during import to Atlas Pro. I consider this an example of a weak data handling facility that is unlikely to see much use. Serious users would always maintain their live data in external data sets and then join or otherwise process them before they are imported into Atlas Pro.

Using Atlas Pro—Building Thematic Maps

As is so often the case, having a defined, moderately thought-out image of the job to be done is valuable to a person setting out to use Atlas Pro to build a map. Most important, of course, is the matter of overall map theme. What message will the intended

FIGURE 1.
Lignite Mines in Four North Dakota Counties. Latitude and longitude locations of each of 134 mines sites were obtained from the State of North Dakota. These point locations were combined with the associated county boundaries that are shipped with Atlas Pro. Underground mines were assigned a value of "1" and strip mines a value of "2" in a spreadsheet program (see Table 1); then provided Atlas Pro as a text file. If this map had to be maintained on an ongoing basis, the Macintosh's publish-and-subscribe capabilities allow attribute data to be provided to Atlas Pro dynamically: when the data were updated on the Excel side, the map would immediately reflect the changes.

Three point symbols were assigned, for the two kinds of mines and for sites about which there was no data, the user can exercise control over the display of points or areas in Atlas Pro for which associated data is not available. Some of the legend in this map had to be built "manually"—what the system provided automatically was not satisfactory.
map try to convey and what is the best means of conveying that message? What will the new map include: how big an area will it cover geographically and what data items are important to the issue at hand? What boundaries, highways, or natural features will serve to make the map's message(s) clearer to the intended reader? What type of data is under consideration and what might be the best choice in its display? What size will the finished map be?

Different data sets are best mapped in varying ways, depending on the nature and extent of the data themselves. Some data are more often viewed as continuous and lend themselves to a dot mapping approach more easily than a choropleth scheme. Nominal point data (descriptive only) can be displayed with different symbols (Figure 1) while ordinal data benefit from symbols that communicate magnitudes to the reader (Figure 2). Ratio data are frequently mapped using the choropleth method. Atlas Pro provides facilities for experimentation using different mapping approaches, scales, patterns, and symbols, but it is worthwhile to consider the overall cartographic impact beforehand.

Boundary files are first used to construct as many layers as the user needs. An Atlas Pro layer is simply a user-defined collection of geographic features (including boundary and data) or text elements that are named and can be managed directly in the program. Many map elements (state boundaries, highways, etc.) could be lumped into a single layer, but experienced users would avoid this practice. Instead, by defining these components separately to Atlas Pro, the user can exercise control over the appearance of

**FIGURE 2.**

*Selected Eastern North Carolina Cities.* Atlas Pro can display point data as symbols (circles here) in sizes proportional to their data value. Several layers have been combined to generate this map: city locations, county boundaries, and interstate highway routes are all part of the standard boundary data provided by SMI. City populations for places with more than 25,000 persons were imported into the program from a database file.

Note that the program cannot produce legends that are useful for this kind of data display, so the guide table in the lower right was imported and superimposed to help readers interpret circles sizes—and their corresponding populations—more readily.
each layer independently in the finished map. Layers and their associated data can be turned "off" or "on" and moved in front of or behind one another. Data can be imported for each layer by making that layer active and using the "Import Attributes" command with its various file format options.

Areas and points on the map can be displayed as labeled or not; individual labels themselves can be moved, deleted, or changed to other typefaces and styles, as needed. The latitude/longitude grid (graticule) can also be displayed with the user specifying line weight, spacing, and associated labeling. One nice Atlas Pro feature is the ability to interrupt drawing of the map on the screen with a click of the mouse. This saves a lot of time and frustration, since a moderately complex map can take 15 seconds to several minutes to redraw on the screen after a small change.

Many of the elements of Atlas Pro's interface are familiar to long-time users of MapMaker. Map components are assembled in fundamentally the same manner as they were in versions 4.0 or 4.5 of Atlas Pro's predecessor. Functions have seemingly been scrambled on the menus when compared with these prior versions, which can be a source of some confusion. Nonetheless, many dialogue boxes are the same; the overall environment turns out to be less different than one anticipates.

The user now is faced with the "real" work: establishing appropriate classes of data (how many and what break points), choosing the best map type, and then getting the map to look exactly as desired. Map types for displaying areal data include choropleth (Figure 3) dot density (Figure 4), and cartogram. The third option, however, does not produce a very useful cartogram—areas are displayed as proportionate to their data value but are not contiguous—and is thus less practical.

Point data can be displayed in two basic ways: by means of point symbols (fixed or graduated in size) and graduated symbols. When choosing point
FIGURE 4. Manufacturing Establishments: Meat Products—1960 and 1989. This is an area map, compiled at the ZIP Code level. It uses dots to represent business establishments. Information for this map started out as a large database: all manufacturing sites by ZIP Code in 4,000+ locations in seven Upper Midwest states. The appropriate number of dots are randomly placed within each ZIP’s area. At a small scale such as this, such an approach works very well and results in a clear display of varying densities. In this example, data applying to manufacturers falling into the specific category “Meat Products” have been broken out and mapped independently. Three kinds of boundary data was combined here: the 4,000 ZIP Codes (boundary files purchased from CACI in Arlington, Virginia), state outlines, and Metropolitan Statistical Areas (MSA) county boundaries that show as dotted lines. The state and MSA boundaries supply a background frame of spatial reference for the map reader. The map maker has control in Atlas Pro over the size of the dots and the number of dots per unit of data.

symbols, the user selects a character from the available type faces on his/her Macintosh. That typically represents a more than ample supply—in most instances an exactly appropriate symbol can be found. In Figure 1, I selected three symbols that can be differentiated and that carry comparable weights when printed: The hollow dot is a Courier character while the star and enhanced asterisk are Zapf Dingbats characters. In addition to choosing from the usual array of Macintosh fonts, SMI ships Atlas Pro with Adobe’s Carta® font for cartographic symbols.

Point symbols can be scaled so that point size—as well as the symbol itself—reflects data magnitude. In the case of Figure 2, all cities could have been represented by stars, with the largest cities earning a 36-point-size star (user controls maximum and minimum point sizes) and the smallest at a 9-point size. Alternatively, several classifications of city size could have been established (25,000–50,000, etc.) and a different symbol (e.g., star, open circle, filled box) assigned to each class.

Only two graduated shapes are available: circles and squares. Figure 2 uses circles where total area of each circle is proportional to the city’s population. Graduated circles for point data can be shown as opaque, transparent (as in Figure 2), or pattern-filled and their overall range of diameters can be controlled to produce visual balance. Graduated symbols cannot be used for areal data. For example, showing ag-
gricultural output within each state as a scaled circle cannot be accomplished directly. (Combining a file of state centroid points with the data and superimposing another layer with state boundaries would get the job done.)

Whether displaying point or area data, classes can be established manually or automatically by Atlas Pro. The capability to create data classes by “n-tiles” (e.g., quartiles) is useful; in other cases, the user will elect to control directly this most important map characteristic. He or she also determines the number of classes to be established and, as well, the associated pattern or symbol for the map. Much of this work is done through legend options, which permit user control over the size and presence of the legend, limited editing options (dollar and percent signs, decimal positions), and display of a frequency histogram as part of the legend, among other things.

Only one attribute can be displayed at a time, which precludes display of two kinds of data. In unusual cases, where the same data attribute is to be displayed on both a point and area basis—for example, funds spent on housing by cities (usually seen as points) as well as at the county (area) level—they can be shown together, but they will both be subject to the same user-defined classes and breakpoints. Except for these less-than-common cases, users must live with the inability to display two or more kinds of data simultaneously.

The customization process is not trivial, but Atlas Pro has many options that help the user progress from a basic map display to a finished product. "Tools," "Patterns," and "Colors" menus provide capabilities for displaying any portion of the map in the desired manner. Among the 100 patterns provided based on the standard Mac set of patterns many are graphically busy and should never be seen in public. More useful to the serious computer cartographer would be a greater number of very plain patterns, including screens of 10, 20, 30 percent gray. A set of drawing tools enables lines and other custom touches to be added.

Text can be entered separately, with user control over typeface, size, and style. Text handling in Atlas Pro is satisfactory, but generally cumbersome. Only one line of text can be entered/edited at a time and the absence of an overall default text style is inconvenient. (Instead of a default, new text is established in the font and style most recently used.) Text and graphic objects can also be imported from outside the system (e.g., a picture of a spreadsheet segment or a diagram) by means of either conventional copying and pasting or via publish and subscribe facilities, and emerge beautifully on the finished product when printed on a LaserWriter or equivalent. Scales can also be added easily and contribute to producing a professional-looking final map.

The map image, once completed within Atlas Pro, can be exported for final touchup to vector-based art programs such as Adobe Illustrator® or Aldus Freehand.® These tools can be used to fill any gaps between a map as Atlas Pro produces it and a desired level of graphic detail or sophistication. Ideally, as many of these last-phase improvements can be done within the mapping software as possible, but on a practical level, some steps are likely to remain before the map is ready for publication.

Additional Atlas Pro Functionality

Areas or points on a map can be selected visually by means of a geographical “Select” command. On the attributes worksheet for the associated layer, the program highlights rows affected by the select command. These points could also be chosen by using a distance-measuring tool (e.g., a 50-mile radius from a given point). Once data records have been selected, a “Summary” command generates a group of basic statistics (See Figure 6; user chooses which ones from a list of a dozen or so) on the currently displayed variable. In this instance, “Total Production . . . ” is the current variable. Detailed data corresponding to one or more variables for the selected rows can also be displayed and exported to the clipboard—for manipulation by spreadsheet or other data handling software—by means of a “Show List” command.

More facilities are provided for further—lower level—selection including groups of areas and points within the scope of the summary and related com-
Additional Atlas Pro Functionality

<table>
<thead>
<tr>
<th>Summary</th>
<th>for</th>
<th>Total</th>
<th>Production (.000 tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases:</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing values:</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean:</td>
<td>11</td>
<td></td>
<td>130</td>
</tr>
<tr>
<td>Standard deviation:</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median:</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum:</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum:</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range:</td>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. Data Summary Based on Map Selection

mands. For example, in a selected state, one can generate statistics only for those cities below 5,000 population. This is a welcome function but is limited to querying a single field at a time and is hobbled by the isolation of one layer's data from another.

Publish and subscribe have been mentioned above and Atlas Pro's support of these System 7 capabilities is a big asset. Especially useful in environments where ongoing updating and transfer of data and maps is an issue, it provides an effective way of moving attribute data into Atlas Pro. Documentation is skimpy in this area, but patience gets it working the way it is intended. On the output side, periodic newsletters and reports (e.g., word processing documents or spreadsheets) can subscribe to a published map, always including the latest version automatically.

Two other powerful Atlas Pro "GIS-type" capabilities include ZIP code aggregation and address matching, both of which have been tested only superficially by this reviewer. Not to imply these are not important program strengths: for GIS-type users and marketing applications, these features are important. Atlas Pro's ZIP functionality alone replaces a good part of GEO Query.*

Atlas Pro also supports what SMI calls "SmartPICTS," bit-mapped images that contain longitude and latitude anchor points. Several examples of these kinds of files are included with the product. These can be imported under existing maps and because of their geographic anchoring, can serve an important visual background function. SmartPICTS files can also be opened up on their own as the basic geography for a map.

Three Basic Limitations

Atlas Pro can, in parts, be described as "excellent," even "near great" as a computer software product. Its flexibility and range of functions are broad and frequently provide users with tools they need to produce desired results. The graphic quality of maps produced in the system can be outstanding. On the other hand, the product is diminished by three general limitations: 1) its overall complexity coupled with inadequate documentation, 2) its persistently high number of small bugs, and 3) several important functional weaknesses. None of these limitations is "fatal," but each is probably experienced—in differing degrees—by all users of the product.

System Organization

Atlas Pro menus are not organized very intuitively. Text formatting sub-commands are found under the "View" menu item. It can be exhausting trying to figure out how to do something, even when you know it can be done. For example, control over display or non-display of the map legend is not found under "Edit Legend"; diligent pursuit locates it under the "Geo Display Options" entry. Looking in the documentation's index under "Legend" is no help whatsoever. Another example for which the index is no help involves putting borders around text entries: you know it's something the program will do, but you just cannot figure how and where it is accomplished. Adding a border must be done independently, by means of the "Object Frame" command: there are no options in the text entry or modification dialogs to do this (even though there were in previous incarnations of the program). There is no index entry under either "Frame" or under the "Text" entry a sub-entry for "Framing." One has to know to look under "Object Frame." If one knew that much, there would not be much of a question in the first place.

126 URISA Journal/Reviews
Documentation

The Atlas Pro documentation is a problem. Any confusions that the user might occasionally experience with this product lie partly in the complexity of Atlas Pro’s menu organization, but mostly in the system’s inadequate documentation. There are two manuals: Learning Atlas Pro and Using Atlas Pro. Neither is short, but both could be improved. The first is not satisfactory as an introduction to the system. Despite a section added at the end that includes a chapter on “The Structure of Atlas Pro,” it lacks a lucid, conceptual underpinning that explains clearly—and with good graphics—how the system is organized and how it works.

The substance of the user’s manual consists of only slightly more than a rehash of menu commands, sub-command by sub-command, presented in that sequence. Indexing is inadequate. There are quite a number of graphics in the documentation, but many of them would be more valuable if they were annotated.

Though Atlas Pro’s manuals are uninspired, SMI is committed to an effort to improve their quality and is working on a revised set which should be available later in 1992.

Operational Matters

As noted above, Atlas Pro is far more stable than its ancestors. Nonetheless, in preparing this review, innumerable minor and several major problems were encountered. At worst, the Mac IIcx and Mac IIfx on which we are running Atlas Pro locked and needed to be rebooted. Much more common were individual problems—many, it seems, involving point display as well as association of patterns with classes of map data—that made life more difficult but not impossible. The full page of changes for version 1.03 leaves one with the feeling that some parts of the product were not tested as fully as others.

One of the most persistent stumbling blocks in Atlas Pro is its inability to thematically map more than one layer at a time. The map features in each layer display—or don’t display, depending on user choice—but only attribute data in the active layer are candidates for appearing on the map. Users can work around this limitation to some degree by using copy and paste facilities but this approach is unsatisfactory because such pasting needs to be done at the very final stage of map production. This means that making revisions to maps built in this manner is cumbersome and time-consuming. From a thematic mapping point of view, the layers in Atlas Pro provide more convenience than functionality.

Attributes must be numeric to qualify for display in Atlas Pro, a fact that impedes easy use of simple nominal data. As mentioned in connection with Figure 1, mines could not be described as “Underground” or “Strip.” Instead they had to be represented as “1” and “2,” respectively. In the map’s legend, either the data ranges associated with the variable can be displayed (in this case, “1” and “2,” but they have no intrinsic meaning), or they can be omitted altogether.

Legends cannot be edited satisfactorily. Options are provided for a leading dollar-sign prefix, a trailing percent sign, and control over the number of decimal places, but this is not enough. Numeric format options such as commas and scaling and user-defined insertion characters are not available for legend display. As can be seen in the sample maps shown here, this can affect the final map appearance. Legends are also unsatisfactory in handling point data. Proportionally sized symbols are not displayed as part of the legend and therefore some other method of informing the reader of the symbols’ values needs to be adopted, as was done by means of a representative data table in Figure 3.

In a broader and more important sense, map legends are one of the weaker links in the system. As mentioned earlier, establishing a proper legend for a specific variable involves choosing data break points, the associated point or areal patterns, and formatting the legend itself (as much as possible) to look its best. This is an important part of getting the mapping process “right” and it must be done for each variable to be mapped. If the series of county-wide maps one is preparing involve area, population density, and family income, then this process has to be done for each. So far, no problem: this careful work is fundamental to the mapping process. The issue is that when you move on to the next variable, the work done is lost: information for only the most recent
legend, with its associated breakpoints, is stored by Atlas Pro for the map as a whole. In other words, a map’s legend cannot be connected in Atlas Pro either to a layer or to a specific data attribute within a layer. This limitation also serves to defeat some of the presumed advantages that might come with having available multiple data sets and multiple fields during the mapping process.

Assigning text attributes and areal patterns is confusing and inconsistent in the system. Clicking on certain text elements reveals their font and style under the “View” menu. In other cases, the “View” menu settings refer to the most recently-selected options. Clicking on a pattern also does not reflect which pattern is the current choice for that area—the most recent selected pattern (which may have nothing whatever to do with the matter at hand) is displayed. In practice these procedures are awkward and confusing.

Only one map can be open at a time in Atlas Pro. It would be a significant improvement to permit several map documents to be usable simultaneously. This is especially important since larger maps can take a considerable time to read and display on the monitor.

There is no ability to group user-added text or graphics objects on the map so they can be moved (or cut and pasted) as a block. This means that tiresome “manual” manipulation is the only means for repositioning objects of this sort, and it must be accomplished one at a time.

For a Mac program that has existed in its various forms for a long time, the fact that parts of its interface are noticeably non-standard is another negative. It remains one of few major programs I know of where the escape key does not cancel a dialog box, and where shift/tab does not move one to the previous field within a dialog box. An “Undo” command appears under the “Edit” menu, but I have yet to encounter a command that can actually be undone. A repeat command would also be an asset for various repetitive editing functions both on the data and map sides of the system.

Charting tools have been added to the system with this release. They allow creation of primitive line, bar, and pie charts. The presence of this facility is wholly puzzling: graphics like this can easily be imported from outside Atlas Pro, where they can be constructed with far more sophistication than they can within. There are other areas in Atlas Pro that clearly warrant attention; facilities such as this seem inconsequential.

**Value vs. Function**

Atlas Pro for the Macintosh retails for $795 but is available at educational and governmental discounts. It represents good value for the substantial investment, especially for those whose ambitions involve production of complex maps at differing scales.

The trade-up fee of $249 from Atlas*MapMaker 4.5 to Atlas Pro, is one of the steepest such charges anyone has seen in the desktop software business. To an existing heavy user of the product, the new version/product offers increased power and the investment is worth it. But barely: This is not a whole new product and it suffers from the flaws already noted.

All computer software products can use functional improvement, Atlas Pro included. If Atlas Pro were more bug-free and if it had even “adequate” documentation, I would be its strongest supporter, awaiting eagerly the improvements we are likely to see in forthcoming versions. Notwithstanding these reservations, I remain an admirer of the product and recommend its use frequently to serious Macintosh users who need to display and analyze spatial data.

For more information about Atlas Pro, contact Strategic Mapping, 4030 Moorpark Ave., Suite 250, San Jose, CA 95117, (408) 985-7400.

William Casey is director of computer-assisted reporting for The Washington Post.
AUTHOR INDEX

Refereed


Croswell, Peter. "Obstacles to GIS Implementation and Guidelines to Increase the Opportunities for Success." Vol. 3, 1: 43.


Egenhofer, Max and Andrew Frank. "Object-Oriented Modeling for GIS." Vol. 4, 2: p. 3.


Features

Bertocci, Roberta and Francesco Brunori, Paolo Canuti, Carlo Garzonio and Sandro Moretti. “Public Safety in Italy: A Study Methodology on Settled Areas.” Vol. 4, 1: p. 86.
Koeppe1, H. W. “CORINE: Toward a European GIS.” Vol. 4, 1: p. 82.
Norman, Margaret. URISA Crossword Puzzle. Vol. 2, 2: p. 73.
Feature Map


Reviews

Books


Monograph


Publications


Research


Software


## SUBJECT INDEX

### REFEREED

#### Analysis and Planning

<table>
<thead>
<tr>
<th>Model</th>
<th>Volume</th>
<th>Lead Author</th>
<th>Hardware</th>
<th>Volume</th>
<th>Lead Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling</td>
<td>2, 2: 35</td>
<td>Watterson</td>
<td>New technologies</td>
<td>3, 1: 6</td>
<td>Chenoweth</td>
</tr>
<tr>
<td>Economic analysis</td>
<td>3, 1: 6</td>
<td>Chenoweth</td>
<td>1, 1: 7</td>
<td>Craig</td>
<td></td>
</tr>
<tr>
<td>Routing</td>
<td>3, 2: 50</td>
<td>Tomaselli</td>
<td>2, 2: 7</td>
<td>de Neufville</td>
<td></td>
</tr>
<tr>
<td>Small area forecasting</td>
<td>3, 2: 3</td>
<td>Lee</td>
<td>Issues</td>
<td>Accuracy/data quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2, 2: 35</td>
<td>Watterson</td>
<td>1, 1: 17</td>
<td>Parent</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 2</td>
<td>Grady</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 32</td>
<td>Rudnicki</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 20</td>
<td>Poe</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 1: 11</td>
<td>Epstein</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 45</td>
<td>Dando</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 2</td>
<td>Grady</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 16</td>
<td>Langran</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 24</td>
<td>Denkers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1, 1: 27</td>
<td>Nyerges</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1, 1: 17</td>
<td>Parent</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 1: 20</td>
<td>Craig</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 2: 25</td>
<td>Nolan</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 10</td>
<td>Frank</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 59</td>
<td>Blinn</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 1: 26</td>
<td>Kjerne</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 2</td>
<td>Grady</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 16</td>
<td>Langran</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 2: 12</td>
<td>Armstrong</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 1: 14</td>
<td>von Meyer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 1: 20</td>
<td>Craig</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 24</td>
<td>Denkers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 10</td>
<td>Frank</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 2: 35</td>
<td>Buyong</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 24</td>
<td>Denkers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 3</td>
<td>Egenhofer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 16</td>
<td>Langran</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 2</td>
<td>Blinn</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 10</td>
<td>Frank</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 59</td>
<td>Blinn</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 1: 26</td>
<td>Kjerne</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 2</td>
<td>Grady</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 16</td>
<td>Langran</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 2: 12</td>
<td>Armstrong</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 1: 14</td>
<td>von Meyer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 1: 20</td>
<td>Craig</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 24</td>
<td>Denkers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 10</td>
<td>Frank</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3, 2: 35</td>
<td>Buyong</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 1: 24</td>
<td>Denkers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 3</td>
<td>Egenhofer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2, 2: 16</td>
<td>Langran</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4, 2: 2</td>
<td>Blinn</td>
<td></td>
</tr>
</tbody>
</table>

### Databases

| Database design/development/ | Volume | Lead Author | Education and training | Volume | Lead Author |
| management | 1, 1: 17 | Parent | 4, 2: 59 | Blinn |
| | 2, 1: 26 | Kjerne | 2, 1: 11 | Epstein |
| | 2, 2: 2 | Grady | 3, 1: 14 | von Meyer |
| | 2, 2: 16 | Langran | 3, 1: 43 | Crosswell |
| | 3, 2: 12 | Armstrong | 4, 1: 32 | Onsrud |
| | 3, 1: 14 | von Meyer | 4, 2: 59 | Blinn |
| | 3, 1: 20 | Craig | 2, 1: 38 | Roitman |
| | 4, 1: 24 | Denkers | 2, 1: 2 | Hintz |
| | 4, 1: 10 | Frank | 3, 1: 14 | von Meyer |
| | 3, 2: 35 | Buyong | 4, 1: 45 | Dando |
| | 4, 1: 24 | Denkers | 1, 1: 17 | Parent |
| | 4, 2: 3 | Egenhofer | 2, 1: 38 | Roitman |
| | 2, 1: 26 | Kjerne | 4, 1: 45 | Dando |
| | 4, 2: 3 | Egenhofer | 1, 1: 7 | Craig |
| | 3, 1: 14 | von Meyer | 4, 1: 32 | Onsrud |
| | 3, 2: 35 | Buyong | 1, 1: 39 | Clapp |
| | 2, 2: 16 | Langran | 4, 2: 47 | Hazelton |
| | 4, 2: 2 | Rudnicki | Level of Government | County government | 4, 2: 32 | Rudnicki |
| | | | Local government | 2, 1: 2 | Hintz |
| | | | Urban government | 2, 2: 35 | Watterson |

---

**Video**

Pierson, Michael. ""A Sampling of Recent GIS Videos."" Vol. 1, 1: p. 88.
<table>
<thead>
<tr>
<th><strong>System Concepts</strong></th>
<th><strong>Volume</strong></th>
<th><strong>Lead Author</strong></th>
<th><strong>Volume</strong></th>
<th><strong>Lead Author</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic modeling</td>
<td>4, 2: 47</td>
<td>Hazelton</td>
<td>3, 2: 77</td>
<td>Van Demark</td>
</tr>
<tr>
<td>Geographic information systems</td>
<td>2, 2: 26</td>
<td>Gimblett</td>
<td>4, 1: 73</td>
<td>Hale</td>
</tr>
<tr>
<td></td>
<td>2, 2: 7</td>
<td>de Neufville</td>
<td>4, 1: 66</td>
<td>Johnson</td>
</tr>
<tr>
<td></td>
<td>3, 2: 3</td>
<td>Lee</td>
<td>2, 2: 48</td>
<td>Robinette</td>
</tr>
<tr>
<td></td>
<td>3, 2: 50</td>
<td>Tomaselli</td>
<td>1, 1: 66</td>
<td>Long</td>
</tr>
<tr>
<td></td>
<td>4, 1: 32</td>
<td>Onsrud</td>
<td>2, 2: 48</td>
<td>Robinette</td>
</tr>
<tr>
<td></td>
<td>4, 1: 45</td>
<td>Dando</td>
<td>3, 2: 70</td>
<td>Koeppel</td>
</tr>
<tr>
<td></td>
<td>4, 2: 59</td>
<td>Blinn</td>
<td>1, 1: 69</td>
<td>Wellar</td>
</tr>
<tr>
<td></td>
<td>4, 2: 3</td>
<td>Egenhofer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land information/record systems</td>
<td>2, 1: 2</td>
<td>Hintz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3, 2: 25</td>
<td>Nolan</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3, 2: 35</td>
<td>Buyong</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4, 2: 20</td>
<td>Poe</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2, 1: 26</td>
<td>Kjerne</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2, 3: 55</td>
<td>Buyong</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3, 1: 14</td>
<td>von Meyer</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1, 1: 27</td>
<td>NyerGES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-purpose cadastres</td>
<td>1, 1: 39</td>
<td>Clapp</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4, 1: 56</td>
<td>Beroggi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-purpose systems</td>
<td>3, 1: 33</td>
<td>Steger</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3, 1: 6</td>
<td>Chenoweth</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3, 1: 20</td>
<td>Craig</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1, 1: 27</td>
<td>NyerGES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-time control</td>
<td>4, 2: 47</td>
<td>Hazelton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk analysis/management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visualizations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FEATURE MAP**

**Analysis and planning**

- Site selection/analysis
  - Volume: 4, 1: 93
  - Lead Author: Crowell

**Applications**

- Environmental assessment/analysis
  - Volume: 3, 1: 85
  - Lead Author: Riggle

- Water resources/systems
  - Volume: 3, 1: 85
  - Lead Author: Riggle

- Site design
  - Volume: 4, 1: 93
  - Lead Author: Crowell

**Databases**

- Database design/development/management
  - Volume: 2, 2: 68
  - Lead Author: Flynn

- TIGER
  - Volume: 2, 1: 74
  - Lead Author: Cooke

- Level of government
  - Volume: 2, 2: 68
  - Lead Author: Flynn

- Rural/small municipalities
  - Volume: 2, 2: 68
  - Lead Author: Flynn

- System concepts
  - Volume: 2, 2: 68
  - Lead Author: Flynn

**FEATURES**

**Analysis and planning**

- Location/spatial analysis
  - Volume: 3, 2: 77
  - Lead Author: Van Demark

**Applications**

- Disaster relief
  - Volume: 4, 1: 73
  - Lead Author: Hale

- Environ. assessment/monitoring
  - Volume: 4, 1: 66
  - Lead Author: Johnson

- Facilities/infrastructure
  - Volume: 2, 2: 48
  - Lead Author: Robinette

- Natural resources
  - Volume: 1, 1: 66
  - Lead Author: Long

- Policy analysis
  - Volume: 2, 2: 48
  - Lead Author: Robinette

**Databases**

- Digital mapping
  - Volume: 2, 1: 71
  - Lead Author: Brown

**Hardware/software technology**

- Distributed systems/networks
  - Volume: 3, 1: 67
  - Lead Author: Zwart

- Micro PCs
  - Volume: 1, 1: 74
  - Lead Author: Scoggings

- Multimedia/hypertext
  - Volume: 2, 2: 42
  - Lead Author: Kindleberger

**Issues**

- Accuracy/data quality
  - Volume: 2, 2: 57
  - Lead Author: Wurman

- Management
  - Volume: 2, 1: 49
  - Lead Author: Zuboff

- Procurement
  - Volume: 3, 2: 67
  - Lead Author: Kindleberger

- History/humor/fiction
  - Volume: 1, 1: 47
  - Lead Author: Johnson

- Downward
  - Volume: 1, 1: 60
  - Lead Author: Burt

- Burt
  - Volume: 2, 1: 47
  - Lead Author: Horwood

- Norm
  - Volume: 2, 2: 73
  - Lead Author: Norman

**Regional concerns**

- Australia
  - Volume: 4, 1: 91
  - Lead Author: Walsh

- Canada
  - Volume: 4, 2: 79
  - Lead Author: Merrick

- Europe
  - Volume: 3, 2: 70
  - Lead Author: Koeppel

- Country
  - Volume: 4, 1: 82
  - Lead Author: Koeppel

- State/provincial
  - Volume: 4, 1: 86
  - Lead Author: Bertocci

**Level of Government**

- County
  - Volume: 2, 1: 63
  - Lead Author: Bates

- Local
  - Volume: 4, 1: 78
  - Lead Author: Keenan

- National
  - Volume: 1, 1: 74
  - Lead Author: Scoggins

- Rural/small municipalities
  - Volume: 2, 2: 42
  - Lead Author: Kindleberger

- State/provincial
  - Volume: 3, 1: 72
  - Lead Author: Livermore

- National
  - Volume: 3, 1: 77
  - Lead Author: Petchenik

- Rural/small municipalities
  - Volume: 4, 1: 78
  - Lead Author: Keenan

- State/provincial
  - Volume: 2, 1: 54
  - Lead Author: Miller

- Rural/small municipalities
  - Volume: 2, 2: 42
  - Lead Author: Kindleberger

- State/provincial
  - Volume: 3, 1: 61
  - Lead Author: Brown

- State/provincial
  - Volume: 4, 1: 78
  - Lead Author: Keenan
<table>
<thead>
<tr>
<th>System Concepts</th>
<th>Volume</th>
<th>Lead Author</th>
<th>Volume</th>
<th>Lead Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM/FM</td>
<td>2, 2: 64</td>
<td>Dueker</td>
<td>3, 2: 81</td>
<td>Roe</td>
</tr>
<tr>
<td>Artificial intelligence</td>
<td>3, 1: 80</td>
<td>Asbury</td>
<td>3, 2: 70</td>
<td>Koeppel</td>
</tr>
<tr>
<td>Geographic information systems</td>
<td>3, 2: 85</td>
<td>Penrose</td>
<td>4, 1: 73</td>
<td>Hale</td>
</tr>
<tr>
<td></td>
<td>1, 3: 50</td>
<td>Antenucci</td>
<td>4, 1: 82</td>
<td>Koeppel</td>
</tr>
<tr>
<td></td>
<td>1, 1: 66</td>
<td>Long</td>
<td>4, 1: 91</td>
<td>Walsh</td>
</tr>
<tr>
<td></td>
<td>2, 1: 54</td>
<td>Miller</td>
<td>3, 2: 99</td>
<td>August</td>
</tr>
<tr>
<td></td>
<td>2, 1: 59</td>
<td>Fletcher</td>
<td>3, 1: 77</td>
<td>Petchenik</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Land information/record systems</td>
<td>Public information systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SDSS</td>
<td></td>
</tr>
</tbody>
</table>