Contents

Conference Preview

6 URISA XXII: Integrating Information and Technology—IT Makes Sense
   Randy Gschwind

In My Opinion

9 GIS/LIS Certification Revisited
   R.R. Chamard

Refereed

11 Modeling Housing and Demographic Diversity at Census Tract Versus Block
   Group Levels of Aggregation
   Lijian Chen

21 Using Object-Oriented Database Management for Feature-Based Geographic
   Information Systems
   Nancy Wiegand and Teresa M. Adams

37 The National Spatial Data Infrastructure: Designing Navigational Strategies
   Steven Frank

Features

57 Using GIS to Predict Erosion Hazard Along Lake Superior
   Carol A. Johnston and James Sales

63 The Evolution of Michigan’s Geospatial Data Infrastructure
   Michael N. Beaulac, Shirley Businski and David Forstat

69 The Ways and Means of Redistricting in Wisconsin
   Ken Strasma and Brenda Haskins

75 The Rising Tide of GIS in Minnesota
   William J. Craig

81 GIS Integrates Water-Quality Data and Remote-Sensed Images
   Joel J. Mouradian, Arthur S. Brooks and David W. Bolgrien

Feature Map

87 Wisconsin Assembly Democratic Caucus Index #1
   Ken Strasma and Brenda Haskins
Reviews

Books

90 Profiting From a Geographic Information System (ed., Gilbert H. Castle)  
Rebecca Somers

95 Toward a Coordinated Spatial Data Infrastructure for the Nation  
Nancy von Meyer

97 Urban Finance Under Siege (eds., Thomas Swartz and Frans Bonello)  
John Owen Beifrens

100 Index

105 Guidelines for Manuscript Submission

Cover: Welcome to Milwaukee! Our cover depicts population density in the greater metropolitan Milwaukee area shown at the census block level. Gray areas are unpopulated; density increases exponentially from greens through yellow, orange and red. The Milwaukee River and tributaries enter Lake Michigan at the harbor.

The map was generated from 1992 Census Bureau TIGER/Line files and Summary Tape File 1B population data using GEODE, Wisconsin's Geodemographic Interface System. GEODE is written in the Arc Macro Language, and provides links to SAS. Developed at the University of Wisconsin-Madison, it is utilized by the Wisconsin Legislature and a number of state agencies. The map was prepared by Jerry Sullivan of the UW-Madison Land Information & Computer Graphics Facility.
Editorial Board

Robert Aangeenbrug
University of South Florida

David Arbot
University of Illinois

Marc Armstrong
University of Iowa

Kate Beard
University of Maine

Richard Brail
Rutgers University

Peter Croswell
PlanGraphics, Inc.

Charles Drinnan
Synercom Technology, Inc.

Earl Epstein
Ohio State University

Steve French
Georgia Institute of Technology

Barry Garner
University of New South Wales

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

Robert Lima
Boshe Institute

John McLaughlin
University of New Brunswick

Gilbert Mitchell
National Geodetic Survey

Joel Morrison
National Mapping Division, USGS

Harlan Onsrud
University of Maine

David Phillips
University of Virginia

Carl Reed
Genasys, Inc.

Heinrich Reinermann
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

Nancy Tosta
U.S. Geological Survey

Barry Wellar
University of Ottawa

Peter Zwart
University of Tasmania

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: Reviewed: Ken Decker, Portland State University, P.O. Box 751, Portland, OR 97207; (503) 464-4042. Features: Warren Ferguson, Plan Graphics, Inc., 10100 Reunion Place, Suite 715, San Antonio, TX; (210) 366-2010. Reviews: Rebecca Somers, Somers-St. Claire, 3157 Babashaw Ct., Fairfax, VA 22031; (703) 204-0033. Maps: Ted Koch, Wisconsin State Cartographer, 160 Science Hall, 550 N. Park St. Madison WI 53706; (608) 262-6852.
The theme of the 1994 URISA International Conference in Milwaukee, Wisconsin is a play on letters, words and acronyms that combine a number of ideas into one simple statement. This theme has been misprinted and misinterpreted, so it’s appropriate to discuss its multiple meanings here in the special conference issue of the URISA Journal.

**Meaning #1:** Integrating information makes sense.

**Meaning #2:** Integrating information is cost-effective ($ & cents).

**Meaning #3:** Integrating technology makes sense.

**Meaning #4:** Integrating technology is cost-effective ($ & cents).

**Meaning #5:** Integrating both information and technology makes sense and is cost-effective!

The theme recognizes our significant progress as information specialists who integrate information and technology in useful ways. But more important, it focuses our attention on the absolutely critical success factors we must carry into the future.

First, Information Technology (IT), and the solutions we propose with it, must make sense. IT must make sense to our managers, to elected officials, to business professionals, to students and to the public we serve; in short, to our customers. IT must make sense in two ways. One is that the technology itself must be accessible to the people who need it. The URISA conference keynoter on Thursday, August 11th—Dr. Michael Dobson—will address the topic of sensitivity to users of the technology. People have to be comfortable with the technology to make IT useful. The other way IT must make sense is in the information and solutions derived. If they make sense, people will use them to make better decisions. This requires our dedication to the ongoing education of our users.

We have spent several decades—this is URISA’s 32nd annual conference—refining and maturing our applications and the resultant information and solutions. The technology has become immeasurably more powerful, and the applications have multiplied to the limits of complexity and imagination. Are we, as a result, providing better information upon which to base decisions? Probably so. We certainly provide more of it and faster. But does our job end here? Is spaceship Earth a safer, cleaner and better place to live? Where is the gap and how can we as information specialists continue to improve our tools and methods to bridge it?

Second, the conference theme proclaims that IT must be cost-effective. The technology itself must be cost-effective to use, and the solutions derived must be cost-effective. We are all enamored with the tools. But even the finest and most expensive tools are worthwhile only if we use them to their full capability to accomplish a desired goal. A shiny new hammer can be a hindrance or an asset, depending on how it is wielded. It can lie in a toolbox or build a house. (It can also blacken a thumb!)

As a collection of tools, IT has a profound impact on how modern organizations operate. The more effectively the tools are used, the more cost-effective they become. But how do we measure this? The conference keynoter on Monday, August 8th—Professor James Quinn—will address gains in operational flexibility, reliability and transaction volumes that are real, but not easily measured with traditional methods.

Our constant, internal IT focus is to strive to use the tools in ways that drive down the costs of achieving the agency mission without sacrificing quality. As a well-organized collection of information, IT can produce internal savings as well as external revenues. The Wednesday, August 10th keynoter—Robert Woods—will address the outward focus toward customer service as the goal of integrating information and technology.

Perhaps, on reflection, the theme is as much a question as it is a statement, as much a dedication to purpose as a pat on the back. We’ve come a long way, but there is far to go. The changes and improvements we make in IT change the way people work. These changes...
are difficult and sometimes even traumatic. As change agents, one of our greatest challenges in implementing new technologies is to understand the difficulties of change and to help people move through them. The global environment is dynamic, as is that of the IT professional. The pressures on the environment are intense, as are those in the IT environment. Let us pledge to improve our tools, methods and applications, and show others how to use them as a synergistic set of solutions to improve our world. IT only makes $ense.

As you read through this conference issue of the URISA Journal, I would ask that you keep the theme in mind. Do these studies advance the state of IT? Do they make $ense to you? Will they make $ense to our customers? We’d be interested in hearing from you. Hope to see you in Milwaukee August 7-11, 1994!
In My Opinion

In this issue . . .
A brief rebuttal to Nancy Obermeyer’s Horwood Critique Article, “Certifying GIS Professionals: Challenges and Alternatives,” published in the Spring ’93 issue of this journal (Vol. 5, No. 1).
Keep those opinions coming!

Consider In My Opinion to be a forum for your thoughts, an opportunity to speak your mind. The only guidelines we suggest are that it:

• Be relevant and well-written.
• Address an issue that is related to URISA.
• Be a topic of current interest or debate, yet not so time-sensitive that the Journal’s publication lead-time would detract from its impact.

We invite you to express your opinion. Please send your submission to Kenneth Dueker, David Moyer, or Bernard Niemann.

The Editors
In My Opinion

GIS/LIS Certification Revisited

R.R. “Sky” Chamard

Professor Obermeyer’s Horwood Critique Prize-winning article recently published in the URISA Journal was both interesting reading and a disappointment. Interesting, for I always enjoy reading about the pros and cons of certification, but it was disappointing that she did not recognize that a certification program already exists for GIS/LIS professionals.

In 1975, the American Society of Photogrammetry initiated its successful Certified Photogrammist program. In March of 1991, the American Society for Photogrammetry and Remote Sensing (ASPRS) expanded the certification program to two new categories: the Certified Mapping Scientist-Remote Sensing, and the Certified Mapping Scientist-GIS/LIS.

The ASPRS recognized, early on, that the GIS community was well on its way to becoming an important profession. The ASPRS was also aware that one of the problems that faced those wanting to hire services in the GIS design and implementation arena was the ability to evaluate those persons that can provide services. Certification provides a medium for evaluation. The ASPRS certification program satisfies most of the concerns addressed by Professor Obermeyer in her award-winning article.

Anyone wanting more information on the program can write ASPRS headquarters at 5410 Grosvenor Lane, Suite 210, Bethesda, MD 20814–2160, or call Sky Chamard, Chair ASPRS Evaluation for Certification Committee (503) 683–2504.

Chamard/URISA Journal 9
In this issue...

Lijian Chen’s offering in this issue will interest users of census data. It addresses the effects of geographic data aggregation upon spatial patterns of housing and demographic diversity. He compares data aggregated by census tract to data aggregated by block group and shows differences in spatial patterns and statistical measures. One of the paper’s referees notes that “it does an excellent job of mixing traditional and not-so-traditional descriptive statistical procedures with the presentation capabilities of an automated mapping system.”

The Nancy Wiegand/Teresa Adams paper presents a persuasive case for moving to an object-oriented paradigm for GIS. Well written and using understandable geographic examples, it provides a readable introduction to the object-oriented approach. It is also rich in insights for those who are already familiar with the approach. A referee comments: “This paper makes a novel contribution as it explains in greater detail than any previous attempt on how to make use of advances in the database community to design next-generation GISs.”

Steven Frank addresses the place of digital geographic data in the evolving world of Internet. The author has done a wonderful job of compiling basic information about network resources and possibilities, and providing a useful bibliography for users to investigate. The greatest contribution of this paper will be to organize this information into a coherent framework relevant to the National Spatial Data Infrastructure.

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 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. 105-106.)

Kenneth J. Duiker
Modeling Housing and Demographic Diversity at Census Tract Versus Block Group Levels of Aggregation

Lijian Chen

Abstract: Using two sets of 1990 census data aggregated at the tract and block group levels respectively, we investigate the diversity of housing and demographic characteristics in the greater Boston area. Our approaches include computations of Gini coefficients of distribution of poor households, location quotients of poor households and less educated adults, and construction of median housing value regression models. In combination with these commonly used conventional socioeconomic analytical approaches, we use GIS technology to present graphically the computational results and plot the error residuals of the housing regression models. The computational results, regression errors, and their spatial patterns vary significantly between the two data aggregation levels. The Gini coefficient analysis reveals a nearly 35 percent greater inequality of distribution of poor households in the greater Boston area at the block group level than the tract. The location quotients calculated using the two sets of data also show significant differences. Plotting the location quotients by area using GIS reveals richer and more complicated spatial patterns of distribution of poor households and less educated population at the block group level. In the analysis of housing value, the model constructed using the block group data reduces the overall explanatory power slightly but contributes to higher significance levels and greater consistency of individual coefficients. Our findings shed light on the importance of selecting appropriate data aggregation levels in the spatial analysis of housing, income, demographic characteristics. We also show the usefulness of a GIS tool in analyzing income distribution and housing markets. It appears that many spatially related urban and regional economic analyses can benefit from the use of GIS.

In this paper we investigate housing and demographic diversity in the greater Boston area using 1990 census data aggregated at two geographic levels: tracts versus block groups. Census data are now available for both tract and block group levels in more easily accessible electronic formats. In the past, most researchers have relied on data aggregated by census tract not because a tract is theorized as an appropriate level to characterize areas differentiated by socioeconomic attributes and demographic compositions, but because more disaggregated data have been difficult to access. Given that 1990 census data for tracts and block groups are equally accessible using today's computing technology, we are interested in whether census data at the block group level are better choices over tract level data. In other words, by applying a number of widely used conventional socioeconomic analytical techniques to census data aggregated at the two different levels, we hope to reveal that important diversities are associated with the differently aggregated data and research results can vary significantly. This is the first objective for this paper.

Our second objective is to explore the utility of GIS in the analysis of housing and demographic characteristics of a neighborhood, either approximated as a census tract or a block group. We use GIS to portray graphically and spatially the analytical outputs generated by the selected traditional approaches of social and economic research. Also, we employ GIS to generate inputs to the analytical processes.

Literature Review and Methodologies

A census tract has been the most frequently used areal aggregation to define a neighborhood or a housing submarket. It has been postulated that a residential neighborhood should be characterized by an area that is large enough to summarize the neighborhood effects that are

Lijian Chen is a PhD candidate in the Department of Urban Studies and Planning at Massachusetts Institute of Technology. Major areas of research include applications of GIS technologies to land use and housing market analysis, and urban and regional economic modeling.
common to small groups of residents, yet small enough to distinguish significant differences in these effects among neighborhoods across the large urban environment (Goodman 1977). However, the universal use of census tract data aggregation by no means implies that a tract is a good representation of a neighborhood. Rather, its popularity is mostly due to data limitations. It is well known that there exist great diversity of housing and demographic characteristics at the tract level. These diversities are lost because of aggregation. For example, according to the 1990 census, Tract 0705 in Boston is characterized by an average of 27 percent households with annual income less than $15,000 and 51 percent of nonwhite population. This tract consists of four block groups which actually display substantial variations in income and racial characteristics. For the block groups, the less-than-$15,000 percentage ranges from 8 to 63, and the nonwhite population from 23 to 86. Such information loss due to aggregation could change one’s research results dramatically. As census data aggregated by block group, typically 20 to 30 percent of the size of a tract, have become readily available, it is desirable to look into the potential of using block group data to represent more homogeneous neighborhoods. It is usually difficult to obtain a lower level of aggregation such as a single block or a parcel group because of the legal confidentiality requirement for compiling census data.

We found only one published work in urban and regional economics literature that explicitly compares the differences of using tract versus block group data. In his study of housing price models for New Haven, Connecticut, Goodman (1977) shows that the block aggregation is a better choice than the tract for constructing housing price models. He finds that measuring the neighborhood variables at the block group level contributes some additional explanatory power to the estimated price models. His results indicate that at the block group level, both the R-square for the model and the significance levels for the coefficients of included socioeconomic variables are improved.

In this paper, we select three approaches to test the effect of data aggregation levels and portray some of these test results using GIS. These three approaches are Gini coefficients, location quotients, and housing value regression models. All three have been commonly employed by researchers in urban and regional studies. We will describe each of these methods briefly below. But first we need to point out one common weakness that seems to be associated with these methods. Although they all have been frequently applied to analyze spatial and location matters, their ways of representing the distribution of spatial and location attributes are either tabulated alphanumeric information or abstract mathematical formulae. These conventional approaches lack an intuitive and easily comprehensible way of portraying location related data and research results. Thus, we have incorporated GIS into experimenting with possible ways of correcting this weakness.

We explain the three approaches in the context of analyzing housing and demographic variables. For those unfamiliar with these approaches, a fuller account can be found in the references cited.

The first test of aggregation levels is to compute a series of Gini coefficients for the distribution of poor (or rich) households. The Gini coefficient is commonly used to measure the inequality of wealth distribution (Gini 1913; Duncan 1955). The Gini coefficient lies between 0 and 1, with 0 representing complete equality and 1 the most unequal distribution. The calculation and meaning of a Gini coefficient can be easily demonstrated with the Lorenz curve of income distribution which gives the cumulative percentage of the total income accruing to the lowest percentiles of the population. Instead of measuring income distribution, we construct Gini coefficients to measure how unevenly tracts or block groups share the proportions of total poor households in the greater Boston area.

The second test is to calculate location quotients of poor households and educational attainment levels by tract and block group respectively, and plot the results using GIS. The location quotient (LQ) is a device frequently used to identify specialization, concentration, or potential of an area in selected employment, industry, or output (Bendavid-Val 1983). Mathematically, the LQ is simply a ratio of ratios, with the top ratio equaling the fractional share of the subject of interest at the local level and the bottom at the regional level. Thus, the LQ can be no less than zero. When the LQ is greater than 1, it means that the local area is more specialized in the subject of interest. The opposite is true when the LQ is less than 1. We use the LQ to identify tracts or block groups that contain higher percentage shares of poor households or less educated population as compared to the greater Boston area as a whole. We then plot and compare the results using GIS mapping capability.

Another test is to construct housing value regression models for the greater Boston area. The regression models allow for testing the effects of several socioeconomic variables simultaneously. The theoretical foundations for these models can be found in the work of Griliches (1971) on hedonic price theory and a rich collection of research by other economists (Ball 1973; Dubin 1990). Simply stated, the market price of a housing unit can be estimated as the weighted sum of a series of values associated with various housing attributes, such as physical structure, accessibility to employment, neighborhood social and economic environments. Following the
spirit of hedonic price theory, we prepare a list of attribute variables based on housing and demographic data from the 1990 census to model median housing values at the tract and block group levels, respectively. We then compare the tract and block group models statistically. Our emphasis is on demonstrating the benefits of using GIS to assist regression analysis. To analyze how well our models predict median housing value for a tract or block group, we plot the regression residuals associated with each observation's location. By doing so, we are able to observe the performance of our models by location and detect any spatial patterns. The approach of mapping residuals was first applied by McCarty (1956) and his colleagues to a study of industrial locations at a national level.

Data and the Study Area

The housing and demographic data used for this research have been extracted from the Summary File Tape 3 of U.S. Census 1990. Geographic coverages for tracts and block groups of the greater Boston area are extracted from the 1990 TIGER/LINE files. The greater Boston area is defined in this paper as consisting of five counties: Essex, Middlesex, Suffolk, Norfolk and Plymouth, as shown in Figure 1. The area contains 827 tracts and 3429 block groups. In 1990, this area was inhabited by 3.8 million of people, 1.43 million households, and 1.5 million housing units.

Figure 1 displays a map of the study area along with some additional income and population information by census tract, serving to acquaint those unfamiliar with the greater Boston area. The shading shows that 23 percent of tracts have $50,000 or higher median household income and 52 percent have between $30,000 and $49,000 income. Most of these wealthy tracts are located in the suburban areas surrounding the city of Boston. The remaining 25 percent of tracts have median income less than $30,000. Most poor tracts can be found in Suffolk County (Boston), as shown in the map inset. A number of other poor tracts with densely clustered population are observable at locations about 20 to 30 miles away from Boston. In this paper, our graphic presentations of data and research results are constrained by the size of the page and the lack of color. However, it is not difficult to imagine that using GIS to display more detail and greater richness of information at the block group level aggregation, in a graphically more attractive manner, is entirely possible.

FIGURE 1. Population Distribution and Median Income by Census Tract (The greater Boston Area, Census 1990; The median income by tract ranges from $5,000 to $103,000. The average population size by tract is 4,571.)
Testing Results

1. Gini coefficients of distribution for poor households

To illustrate how a Gini coefficient of distribution of poor households is calculated, we define poor households as those households with annual income less than $15,000 and calculate the Gini coefficient using tract data. First, we compute the number of the poor households for each tract as a percentage of the total poor households in the entire region and the total households in each tract as a percentage of the total number of households for the region. Second, we rank all tracts in ascending order according to their percentage shares of poor households. Then, as depicted in Figure 2, beginning with the tract having the least share of the poor, we plot each tract’s cumulative percentage share of poor households along the vertical axis and its cumulative percentage share of the total households along the horizontal axis. By doing so, we obtain a curve, denoted by “Tract Data,” that is convex away from the diagonal line. The curve is widely known as a Lorenz curve. The closer to the diagonal that a Lorenz curve is, the more equalitarian is the distribution. The diagonal line represents complete equality. Thus, the “Block Group Data” curve, the Lorenz curve derived from the block group data, represents greater inequality than the tract curve. The Gini coefficient equals twice the area between the curve and the diagonal. The coefficient lies between 0 and 1. In the case of distribution of poor households, a zero Gini coefficient would mean that every subarea (i.e., tract or block group) shares an equal portion of the total poor households in the greater Boston area. A higher Gini coefficient means that poor households are more unevenly distributed across the region. The Gini coefficients for the block group and tract curves are respectively 0.318 and 0.236.

The 1990 census reports 25 household income categories, among which 12 are below $35,000, as indicated in Table 1, and 13 are greater than and equal to $35,000. We use each of the 12 categories in turn as our benchmark to define poor households. Then we calculate the Gini coefficients of distribution for each of these differently defined poor households by tract and block group, respectively. The results, as presented in Table 1, show that the Gini coefficients computed using the block group data are consistently higher than those using the tract data. The differences are in the neighborhood of 35 percent. These results should make one be cautious about selecting appropriate data aggregation levels.

2. Location quotients of poor households and less educated population

In this test of data aggregation levels, we compute the LQs to examine each local area’s relative share of poor households and less educated population as compared to those of the whole region. We define poor households as those with annual income less than $15,000 and less educated population as those with less than high school degrees. To illustrate, one formula of computing an LQ for poor households for a tract is:

\[
LQ = \frac{TP_i}{TTi} \times \frac{TRP}{RT}
\]

where

- \(TP_i\) = total number of poor households in Tract i,
- \(TTi\) = total number of households in Tract i,
- \(TRP\) = total number of poor households in the greater Boston area,
- \(RT\) = total number of households in the greater Boston area.

The location quotients calculated using the tract and block group data separately are shown in Table 2. For the convenience of comparison, we devise the two distributions of LQ values at <1.0 and >=1.0, with 1.0 being the value where the percentage of the poor or less educated population is the same as that of the region. Two differences between the tract and block group calculations are observable from this table. First, for both LQs, the proportion of census tracts that share a higher percentage of the poor and less educated than the regional counterparts is about 4 percent higher than that of block...
### TABLE 1. Comparison of Gini Coefficients of Distribution of Poor Households Between the Tract and Block Group Data Aggregations

<table>
<thead>
<tr>
<th>Income Category</th>
<th>Number of Households</th>
<th>As % of Total Households</th>
<th>Cumulative % of Households</th>
<th>By Tract</th>
<th>By Block Group</th>
<th>% of Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; $5,000</td>
<td>58,143</td>
<td>4.1</td>
<td>4.1</td>
<td>0.352</td>
<td>0.477</td>
<td>35.5</td>
</tr>
<tr>
<td>5,000 to 9,999</td>
<td>118,294</td>
<td>8.4</td>
<td>12.5</td>
<td>0.289</td>
<td>0.390</td>
<td>34.9</td>
</tr>
<tr>
<td>10,000 to 12,499</td>
<td>44,696</td>
<td>3.2</td>
<td>15.7</td>
<td>0.260</td>
<td>0.349</td>
<td>34.2</td>
</tr>
<tr>
<td>12,500 to 14,999</td>
<td>36,583</td>
<td>2.6</td>
<td>18.3</td>
<td>0.236</td>
<td>0.318</td>
<td>34.7</td>
</tr>
<tr>
<td>15,000 to 17,499</td>
<td>43,911</td>
<td>3.1</td>
<td>21.4</td>
<td>0.214</td>
<td>0.289</td>
<td>35.0</td>
</tr>
<tr>
<td>17,500 to 19,999</td>
<td>40,398</td>
<td>2.9</td>
<td>24.3</td>
<td>0.198</td>
<td>0.265</td>
<td>33.8</td>
</tr>
<tr>
<td>20,000 to 22,499</td>
<td>49,075</td>
<td>3.5</td>
<td>27.7</td>
<td>0.181</td>
<td>0.240</td>
<td>32.6</td>
</tr>
<tr>
<td>22,500 to 24,999</td>
<td>39,192</td>
<td>2.8</td>
<td>30.5</td>
<td>0.167</td>
<td>0.223</td>
<td>33.5</td>
</tr>
<tr>
<td>25,000 to 27,499</td>
<td>50,112</td>
<td>3.6</td>
<td>34.1</td>
<td>0.152</td>
<td>0.202</td>
<td>32.9</td>
</tr>
<tr>
<td>27,500 to 29,999</td>
<td>40,167</td>
<td>2.8</td>
<td>36.9</td>
<td>0.141</td>
<td>0.187</td>
<td>32.6</td>
</tr>
<tr>
<td>30,000 to 32,499</td>
<td>54,242</td>
<td>3.8</td>
<td>40.8</td>
<td>0.127</td>
<td>0.169</td>
<td>33.1</td>
</tr>
<tr>
<td>32,500 to 34,999</td>
<td>39,440</td>
<td>2.8</td>
<td>43.6</td>
<td>0.118</td>
<td>0.156</td>
<td>32.2</td>
</tr>
<tr>
<td>&gt;= $35,000</td>
<td>795,985</td>
<td>56.4</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1,410,238</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: calculated based on US Census 1990.

### TABLE 2. Location Quotients of Poor Household and Less Educated Adult by Tract and Block Group

<table>
<thead>
<tr>
<th>Location Quotient</th>
<th>Number of Tracts</th>
<th>As % of the Total</th>
<th>Number of Block Groups</th>
<th>As % of the Total</th>
<th>Number of Tracts</th>
<th>As % of the Total</th>
<th>Number of Block Groups</th>
<th>As % of the Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>zero</td>
<td>4</td>
<td>0.5</td>
<td>141</td>
<td>4.2</td>
<td>11</td>
<td>1.3</td>
<td>145</td>
<td>4.2</td>
</tr>
<tr>
<td>&lt;1.0</td>
<td>453</td>
<td>55.9</td>
<td>1,939</td>
<td>57.4</td>
<td>437</td>
<td>53.0</td>
<td>1,810</td>
<td>52.8</td>
</tr>
<tr>
<td>&gt;=1.0</td>
<td>353</td>
<td>43.6</td>
<td>1,298</td>
<td>38.4</td>
<td>377</td>
<td>45.7</td>
<td>1,471</td>
<td>42.9</td>
</tr>
<tr>
<td>Total</td>
<td>810</td>
<td>100.0</td>
<td>3,378</td>
<td>100.0</td>
<td>825</td>
<td>100.0</td>
<td>3,426</td>
<td>100.0</td>
</tr>
<tr>
<td>Missing</td>
<td>14</td>
<td></td>
<td>48</td>
<td></td>
<td>2</td>
<td></td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Source: calculated based on US Census 1990.

Note: Poor households are defined as those with annual income less that $15,000; less educated adults are those 25-year old or above with education attainment less than high school graduates.
groups. Second, the poor households are concentrated in fewer tracts or block groups than the less educated adults. These comparisons again indicate that there are important differences between calculations based on different aggregation levels. However, the descriptive statistics in the table tell us little about where these poor and less educated areas are and how they relate to each other spatially. Using a GIS tool, we are able to provide additional important evidences for concerns about the choice of data aggregation levels.

Figure 3 shows the spatial distribution of location quotients for the poor households. It is not hard to see that the location quotients plotted by tract can be very misleading. For example, the tract map shows that few areas have LQ equal to zero, indicating those localities with no poor households. In contrast, the block group map shows that there are hundreds of this type of block groups scattering across the region. This difference can be attributed to the way that the LQ is calculated. It is much more difficult for a tract to have a zero LQ because each of the block groups within this tract needs to have a zero LQ in the first place. It might not be very difficult to note that in Table 2, it is unlikely that all those 141 block groups with zero LQ of poor households would all fall within the three tracts of the same characteristics. But it would be very difficult to gain a good understanding how these block groups differ from those tracts in the real space without representing these locations graphically using GIS. The comparison of the two maps indicates that the differences of LQ for poor households between displaying the tract and block group data can be very dramatic. The same conclusion can be drawn for the LQs of less educated adults, as shown in Table 2.

The plotting of LQs also allows us to understand where those tracts and block groups with higher proportions of poor households and less educated population than the regional average are located, how these localities relate to each other (clustering around a number of centers or spreading all over), and whether some of these localities cluster to form a critical geographic mass that spans across administrative boundaries, such as those of counties and towns. Our correct answers to these questions would be of important value for further research design and many policy and program initiatives.

3. Housing value regression models

We wish to estimate linear regression models to explain the variations of median housing values by census tract and block group respectively for the greater Boston area. Our goal is two-fold. First, we hope to explore how census data aggregated at different levels affect the results of regression analysis of housing value. Second, and more importantly, we attempt to illustrate that GIS can be a very useful tool in the process of formulating research questions, constructing models, and interpret-

---

**FIGURE 3.** Location Quotients for Poor Households: Census Track vs. Block Group (The greater Boston Area, Census 1990; Numbers next to the legend categories are sample distributions for block group and tract respectively.)

- **Map by Block Group**
  - Total number of block groups: 3426

- **Map by Census Tract**
  - Total number of tracts: 824

---
### TABLE 3. List of Variables Included in theEstimatedHousing Value Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOUSING_VALUE</td>
<td>Median value of owner-occupied housing units (dependent variable)</td>
</tr>
<tr>
<td>LOW_INCOME%</td>
<td>Percentage of households with annual income less than $25,000</td>
</tr>
<tr>
<td>HIGH_EDUCATION%</td>
<td>Percentage of adults (&gt;=25 years) with Bachelor's degree or higher</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>Distance (in miles) from the CBD, Boston (the CBD is defined as the center point of the block group where Boston Common is located)</td>
</tr>
</tbody>
</table>

Possible independent variables to explain the areal variations of median housing value. However, many of these possible variables are highly correlated with each other. For example, variables measuring educational attainment are highly correlated with income; the percentage of non-white population with the proportion of low-income households; and the distance away from the CBD with the population density-related variables. To avoid the potential problem of multicollinearity among the independent variables and to keep regression models simple, after numerous model runs, we developed a fairly good-fitting, simple model with only four variables, as described in Table 3. Among these variables, distance is the only variable that is not directly derived from the census data. We compute distance from the geographic center of each block group or tract to the CBD of Boston based on the geographic coverages of block group and tract converted from the TIGER/LINE files. The computation of distance was completed by using GIS built-in functionalities to first derive the coordinates of centroid for each block group and tract respectively, and then calculate distance of each centroid to the assumed CBD coordinates.

### TABLE 4. Housing Value Models for the Greater Boston Area: Tract Versus Block Group Data, Census 1990
(dependent variable = HOUSING_VALUE)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Parameter Estimates</th>
<th>Pearson Correlation Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficients</td>
<td>Standardized Coefficients</td>
</tr>
<tr>
<td>Using Census Tract Data:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>94,790</td>
<td>11.829</td>
</tr>
<tr>
<td>LOW_INCOME%</td>
<td>78</td>
<td>0.015</td>
</tr>
<tr>
<td>HIGH_EDUCATION%</td>
<td>3,377</td>
<td>0.811</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>-586</td>
<td>-0.077</td>
</tr>
<tr>
<td># of observations</td>
<td>784</td>
<td></td>
</tr>
<tr>
<td>F-statistics</td>
<td>540.16</td>
<td></td>
</tr>
</tbody>
</table>

R-square: 0.675
Adj. R-square: 0.674

Using Block Group Data:

| Constant             | 110,256             | 30.042                         |              | -0.385       | 0.775       | 0.190          | 0.150          |
| LOW_INCOME%          | -101                | -0.021                         | -1.586       |              |            |                |                |
| HIGH_EDUCATION%      | 3,076               | 0.752                          | 56.670       |              |            |                |                |
| DISTANCE             | -621                | -0.080                         | -6.893       |              |            |                |                |
| # of observations    | 3,172               |                                |              | -0.385       | 0.775       | 0.190          | 0.150          |
| F-statistics         | 1,623.92            |                                |              |              |            |                |                |

R-square: 0.606
Adj. R-square: 0.606
change in each of the independent variables on the dependent variable. The standardized coefficients in the second column compare the importance of the independent variables in determining the median housing values. The most important determinant of housing values is the percentage of adult population with high education degrees. Also presented in the table are the correlation coefficients between those variables included in the models.

All estimated coefficients are significant at the 1 percent level except for those associated with the variables of percentage of low income households. In the block group model, as expected, the sign on this income variable is negative and nearly statistically significant at the 10 percent level. However, the same variable has a wrong sign and is statistically insignificant in the tract model. The significance levels for all independent variables but LOW_INCOME% are considerably higher for the block group data than for the tract. Therefore, our regression results seem to suggest that using the block group data contribute to higher significant level and greater consistency of individual coefficients. In terms of overall explanatory power, the tract model outperforms the block group slightly. The adjusted R-square declines from 0.675 for the tract model to 0.606 for the block group model. This decline is not surprising because it is generally true that when data are aggregated at higher levels, the variance for each regression variable is reduced. Therefore, models estimated for higher aggregation levels usually achieve higher R-square values, all other things equal.

The magnitude of impact of the independent variables on the median housing value differs between the tract and the block group models. According to the tract model, if two tracts that differ by one percentage point in proportion of adult population with bachelor's or higher degree, the median housing value of the tract would differ by $3,377. For the block group model, the difference would be as high as $3,076 for each percentage point difference in HIGH_EDUCATION% between two block groups. In contrast, the distance variables show a slightly stronger negative impact on the housing values in the block group model than in the tract. For each mile away from the CBD (Boston), median housing value for a block group would drop by an estimated $621, noticeably higher than a $586 coefficient for a tract. The ways in which the distance variables behave in these models very much reflect the spirit of residential location theory formulated by Alonso (1964) and Muth (1969). This theory states that the market value of a standardized housing bundle declines with distance from the city center.

Using GIS, we are able to make further comparison between tract and block group aggregations by plotting the standardized error residuals for the two models, as shown in Figure 4. Tracts and block groups are shaded on the basis of the size of their corresponding standardized residuals estimated by the models. The following differences between the two maps of residuals are worth noting.

First, the block groups where median housing values are overestimated or underestimated by more than one unit of standardized error residual equal slightly over 20 percent of the total number of block groups. For the tract model, overestimations and underestimations occur in only less than 17 percent of the total tracts. One might not consider that there exist significant differences in the numerical distributions of error residual between the tract and block group models. However, it is readily observable that the patterns of spatial distributions of overestimations and underestimations are very different, as seen in Figure 4. On the block group map, many overestimations and underestimations occur in the peripheral areas of the region. In contrast, on the tract map, the areas of large estimation errors are mostly situated in the areas close to the center of the region. It is clear from these comparisons that the display of the residuals is useful in understanding the behaviors of the models and interpreting the model results.

Second, in a dozen tracts situated west of the CBD (Boston), and forming a roughly vertical band, the median housing values are consistently underestimated more than one unit of standardized errors by the tract model. The use of GIS to portray the error residuals by location lends us an easy and intuitive way of identifying areas where the estimated models seem to fail apart. It enables us to quickly focus our attention on the problematic areas, initiate extra efforts to investigate other possible causes for the problems, and to reformulate research hypotheses. We soon found out that these tracts form the six or seven most wealthy towns in the greater Boston area. Our tract model predicts poorly for these towns because many factors that are unique to these locations have not been considered in the models, including zoning, minimum lot size, and public service choices and quality. It is also readily apparent from the maps that the block group model is more successful than the tract model in explaining the change of housing values for those wealthy towns.

Third, an area where housing values are overestimated using the tract data may change to one with housing being undervalued when using the block group data, and vice versa. For example, the areas that are situated on the far left side of the region show overestimated median housing values on the tract map and underestimated value on the block group map. Thus, it seems that models estimated using data aggregated at different levels generate different predictions.
FIGURE 4. Error Residuals of Housing Value Models: Census Track vs. Block Group (The greater Boston Area, Census 1990; Numbers next to the legend categories are sample distributions for block group and tract respectively.)

Conclusions

We have used two sets of census data aggregated at the tract and block group levels, respectively, to examine the diversity of housing and population characteristics. Three conventional analytical techniques—Gini coefficient, location quotient and regression analysis—have been used to test the differences between using the tract and block group data. We find that different data aggregations can lead to significantly different research results and conclusions. The Gini coefficient analysis reveals a nearly 35 percent greater inequality of distribution of poor households in the greater Boston area at the block group level compared to the tract. The location quotients calculated using the tract and block group data also show alarming differences. By plotting the location quotients by area using GIS, we find that much richer and more complicated spatial patterns of distribution of poor households and less educated population are revealed at the block group level. The analysis of housing value regression models sheds additional light on the important dissimilarities in different levels of data aggregation. The model constructed using the block group data reduces the overall explanatory power slightly, but contributes to higher significance levels and greater consistency of individual coefficients.

From these three sets of analysis, it can be concluded that research results are very sensitive to the definition of the areal entities. Also, it can be concluded that the block group data seem to be most valuable when one’s research goal is to reveal and take into consideration as much of the diversity within a locality as possible. The plotting of model error residuals appears to be a very instructive way of examining the performance of regression models by location. It provides a useful basis for identifying ways to further improve regression models.

This paper also shows that GIS technology can make useful value-added contributions to the conventional analysis of income distribution and housing market by examining them spatially and by improving the construction of regression model that uses areas as its unit of analysis. We are currently exploring utility of some high-end GIS functionalities. For example, the GIS network capability can be used to measure the distance variable in a more realistic manner, measured according to the actual street and highway network instead of based on the straight line distance assumption. Also, the GIS union and overlay capabilities can be used to obtain more homogenous housing submarkets for the purpose of improving housing hedonic price models. We believe
that many other spatially related urban and regional economic analyses can benefit from the use of GIS.

Acknowledgments

I am grateful to Professors Lyna L. Wiggins, Joseph Ferreira Jr., Frank Levy and Briton Harris for their inspiration, and many generous and valuable criticisms and suggestions on this paper. I would also like to thank William Craig, David L. Phillips and the anonymous reviewers for their numerous helpful and constructive comments and suggestions for revising this paper. Finally, I would like to gratefully acknowledge the computing service support of the Computer Research Laboratories of the School of Architecture and Planning at Massachusetts Institute of Technology.

References


Using Object-Oriented Database Management for Feature-Based Geographic Information Systems

Nancy Wiegand and Teresa M. Adams

Abstract: The potential for using a Database Management System (DBMS) as a framework for a GIS has been recognized. Current relational database systems lack the modeling power and extensibility needed for complex objects and operations such as are used in geographic applications. New DBMS with object-oriented modeling and extensibility form a promising tool to fulfill needed requirements for building a GIS. In addition, new object-centered GIS models have been proposed. By illustrating the actual constructs of an Object-Oriented Database Management System (OODBMS), the potential implementation of an object-centered GIS is shown. The benefits of object-oriented modeling versus relational modeling are illustrated in a translation of an application from the relational model to a feature-based OODBMS model. Also, other OODBMS features and extensibility facilitating GIS are described.

Most current geographic information systems do not have the benefit of full-fledged Database Management System (DBMS) features. Those that do are based on the relational model which does not easily handle the complexity of geographic data. The premises of this paper are that a better GIS can be built by 1) using a DBMS as a base for the system, 2) using object-oriented modeling, and 3) having an extensible system. This paper illustrates these ideas using an extensible Object-Oriented Database Management System (OODBMS).

Many GIS are written as special spatial application systems without useful DBMS features such as physical data independence, query optimization, concurrency control, recovery, authorization, consistency constraints and dynamic schema changes (Frank 1988; Haas 1991). Also, using a DBMS as a framework for a GIS frees programmers from building and maintaining much of the complexity of a large software system (Frank 1988; Egenhofer 1992).

Next-generation DBMSs are being designed to be object-oriented and extensible. The modeling features of object-oriented DBMS are more suitable than the standard relational model for GIS data because the logical and physical structures more directly capture original semantics. Also, using a next-generation DBMS with extensible features enables a design to be tailored to an application.

Many researchers have discussed the potential advantages of using a DBMS or the object-oriented model for a GIS (Frank 1988; Kjærne 1990; Worboys 1990; Haas 1991; Williamson 1991; Egenhofer 1992; Newell 1992; Viljbrief 1992; Adams 1993; Crosbie 1993; Milne 1993). A survey of database support for GIS can be found in Medeiros (1994). It has been proposed that an object-oriented model in which data are organized by complex objects is better than the relational model because related data can be kept together and relationships can be directly modeled. This paper illustrates those proposals by using the constructs of an actual OODBMS for a GIS. The OODBMS referred to here is the Object Database and Environment (ODE) (Agrawal 1991).

Furthermore, the GIS model presented in this paper is the cartographic feature-based object model (Guptill 1990) which focuses on entities rather than on spatial representations as is currently done. The focus on entities or objects lends itself to an object-oriented implementation. This paper integrates database management technology and object-oriented modeling with the feature-based GIS concept, the primary focus being on OODBMS technology.

We begin with a summary of the limitations of current DBMS for a GIS and describe related work including advancements in DBMS technology. The ODE
OODBMS is introduced, and the feature-based object model is presented followed by an overview of its implementation in an OODBMS. After this, data in a current relational GIS are transposed into an object-based model, and the potential implementation of the new model using ODE is described. The advantages of object-oriented modeling compared to relational modeling are illustrated. Finally, other OODBMS features that can meet the needs of a GIS are discussed. The summary includes a discussion of the extensibility of ODE for a GIS.

Limitations of Current DBMS for GIS

The limitations of current DBMS for new complex applications include limitations of relational modeling and lack of extensibility. These limitations led to the development of the next generation of DBMSs.

To review, in a relational DBMS, a table is created for each entity type. A row in a table corresponds to an entity, and columns contain the attribute values. However, a relational schema ends up with many additional tables because an attribute is restricted to being a simple built-in type, an attribute may not be a set of values, and relationships are also modeled using tables. A later section of this paper elaborates on this using examples. GIS spatial and non-spatial data are complex enough that spreading entities and attributes into numerous tables is undesirable. Object-oriented modeling keeps related data together. Currently, GIS systems that have object-oriented modeling are often built on top of a relational model (e.g., Crosbie 1993). Because the type of model is changed between levels of the system, design alternatives may be restricted and performance is possibly reduced.

Current DBMS also lack extensibility to provide for special application needs. Examples of extensible features not in traditional systems are provisions for the user to add new data types and methods, add user-defined code, design new storage methods, and have access to standard packages. A GIS needs extensibility, for instance, to provide seamless, scaleless maps (Guptill 1988). This is an important goal in GIS (Guptill 1989; Becker 1991; van Oosterom 1991). Extensibility, object-orientation, and the new geographic data models based on a set of feature objects (e.g., the feature-based model) provide the needed framework for a scaleless, seamless database. Scalelessness is provided by using abstract data types and user-defined indexes, and the scaleless requirement is met by using multiple representations, user-defined operators and rules (Guptill 1989).

Haas (1991) offers the following extensible DBMS characteristics needed by GIS:

- Ability to have new set operators as part of the query language, e.g., overlay.
- Ability to add new access methods such as a Grid-File (Nievergelt 1984) or R-tree (Guttman 1984).
- Support of new data storage methods so that, for example, a map could be stored as a quadtree or an image could be stored on optical disk.
- Extensibility of the optimizer to allow optimizing with different data types and operators.
- Ability to access user-defined code and standard packages.

Selected Related Work

Next Generation DBMS

An active area of research in database management systems is to provide a next generation of systems with better modeling and extensibility. These systems are referred to as object-oriented or extended relational systems. Examples from (Cattell 1991) are POSTGRES, Starburst, O2, ObjectStore and GemStone.

These new systems evolved in two main ways. O2, ObjectStore, Gemstone, and ODE (Agrawal 1991) were designed by extending existing object-oriented programming languages to have database features such as persistence, concurrency control, and queries. Thus, they are referred to as "object-oriented." On the other hand, Starburst and POSTGRES extended existing relational database systems to have object-oriented programming language capabilities and are thus referred to as "extensible" or "extended relational" systems. The latter term will be used in the remainder of this paper to avoid confusion with the term "extensibility" which can be applied to both relational and object-oriented systems. Object-oriented and extended relational systems are growing toward each other and will probably be indistinguishable in the future (Carey 1990b). (Kim 1993) also foresees the eventual unification of relational and object-oriented technologies. Other DBMS architectures do not fit into the above two categories. For example, EXODUS (Carey 1990a) is a database system generator. It is not a DBMS itself but allows a database implementor to generate a customized DBMS. Extended relational, object-oriented, and DBMS generator systems all have various degrees of extensibility (Carey 1990b). Designers of extended relational systems have explicitly identified pieces of the system to which they will allow the user to add code. On the other hand, object-oriented systems have some natural amount of extensibility since they are usually accessed via a programming language. In this paper an object-oriented DBMS is used rather than an extended relational system. The object-centered GIS we are modeling benefits from the object-oriented modeling and features of an OODBMS espe-
cially because of the ease of storing relationships in an OODBMS.

**Object-Oriented Modeling and Implementation**

Smith and Zdonik (1987) looked at modeling, though not specifically for a GIS. They compared the relational modeling of their electronic documents application implemented in INGRES (Stonebraker 1976) to object-oriented modeling in Encore, an OODBMS. Their conclusion is similar to ours in that they found complex objects were more suited to object-oriented modeling rather than being divided into many normalized relations.

Extensible and object-oriented DBMS show promise for designing and building GIS software and providing better modeling capabilities for applications. (Egenhofer 1992) discusses this and then illustrates the benefits of object-oriented modeling for geographical applications. (Kjerme 1990) has illustrated the advantages of object-oriented modeling using Smalltalk-80, in particular being able to encapsulate methods within an object for dynamically re-calculating locational values. Since Smalltalk-80 does not have persistence, they speculated on the advantages of using a database system, in particular, an OODBMS. In this paper, the use of an actual OODBMS will be illustrated.

A project by Williamson (1991) implemented an object-oriented GIS for an image system using the OODBMS Vbase by Ontologic, Inc. We are taking a different approach by considering a feature-based GIS model and a vector system. Haas (1991) used an extended relational DBMS, Starburst, to model GIS data. Their model used relational tables with added extensible features such as user-defined types, calling a user function from SQL, and writing a spatial join operator. In this paper an object-oriented model and a feature-based approach is used instead of the extended relational model, but like (Haas 1991), extensibility is considered. In the Sequoia Benchmark project (Stonebraker 1993) tested several systems, including the extended relational system POSTGRES (Stonebraker 1990), to load and execute queries on large, image and vector, earth-science data sets. GEO++ (Vijbrieff 1992) is a prototype GIS built as a front-end to POSTGRES. The extensibility of POSTGRES was used to define geometric abstract data types and add spatial operations, and C++ was used to provide inheritance in spatial types. Our project models an entirely feature-based GIS with a focus on relationships using object-oriented modeling within an OODBMS context.

(Crosbie 1993) believes the future of GIS involves object-orientation, and he integrated object-oriented design into an existing standard relational system to form a descendant of TIGRIS (Herring 1987; Herring 1992).

However, using an OODBMS directly avoids translation routines. (Milne 1993) modeled geometric objects using object-orientation and then did performance tests retrieving complex objects using ONTOSS, an OODBMS, and compared the results to relational DBMSs. As seen later, our geometric modeling differs from theirs, and our focus is on a feature-based design for an entire GIS. Smallworld GIS (Newell 1992) comes the closest to our work in its implementation of an object-centered model using an object-oriented programming language. However, Newell did not use a DBMS and worked on top of a tabular system.

**An Object-Oriented DBMS**

**Object-Oriented Programming and C++**

Although object-oriented programming is not well defined in terminology (Nelson 1991), the following equation covers the basic points: "object-oriented = objects + classes + inheritance" (Wegner 1987). An object can be defined as containing values for variables and being manipulated by a set of methods. A class is a description of similar objects.

Inheritance is one of the main characteristics of object-oriented languages. A subclass (derived class) inherits variables and methods of its superclass (base class). Inheritance allows type/subtype relationships. Subtyping allows specialization for the is-a relationship; if class B inherits from class A then B is_a A. As a simple example, the class "toll road" could inherit variables and methods from the superclass road since a toll road is a road. Multiple inheritance allows a class hierarchy to be built for subtyping. For a more complete discussion of object-oriented concepts see, for example, (Nierstras 1989).

Object-oriented features have been added to the C language to form the object-oriented programming language C++ (Stroustrup 1986). The key concept in C++ is the class which is a user-defined type. Classes allow the user to design new objects with encapsulated methods to closely match concepts in the application. In very simple terms, for someone familiar with C's structures or Pascal's record types, a class is similar except that methods (called member functions) are included in the class definition. Furthermore, individual variables (data members) and the member functions can be declared public or private allowing data hiding. Values, data members, and member functions can be inherited.

The pointer data type of programming languages has new significance for an object-oriented database system in that it can be used to model relationships between objects. In this paper, the term "reference pointer" is used.
for a pointer indicating a relationship. For example, an object can have a reference pointer to a subobject to indicate a "composed_of" relationship.

C++ forms a good basis for a database programming language but lacks persistence and other necessary database features.

**ODE and O++**

ODE is an object-oriented DBMS made available for this work by AT&T Bell Laboratories (Agrawal 1991). The language of ODE, O++, extends the C++ object-oriented programming language to suit database needs. Version 2.0 of O++ includes persistence, clusters, iterators, versions, and transactions. Persistence allows objects to survive between program executions. Clusters are physical groupings of objects in a class, and iterators allow set-level query processing of the class objects. Features available in future releases of O++ include sets, subclusters, constraints, and triggers. OODBMSs provide a set construct to directly model multi-valued attributes. This results in a significant modeling advantage over the traditional relational model as will be shown later.

Writing an application in ODE involves programming in C++ combined with O++'s database constructs. The C++ class is the basis of the object model of O++. Defining a C++ class allows the database programmer to design abstract data types (objects and methods) needed for a particular application. This is one way in which O++ is extensible. That is, a programmer defines a class for each main entity type in the application (e.g., for wells, bridges, and points). Each class definition contains attributes for data values, reference pointer attributes for modeling relationships, and methods for operations on the class objects.

**An Object-Based GIS**

**The Feature-Based Model**

GIS models can be divided into object-based and layer-based models (Goodchild 1992). Traditionally, GISs have been implemented with the layer-based model, but new object-based models are being developed (Herring 1987; Newell 1992). The layer-based approach focuses on layers of spatial components of maps, e.g., a layer of line segments denoting roads. In the layer-based approach, non-spatial data are attached to the spatial representations. On the other hand, object models focus on mapped entities, such as Highway 151, with the spatial representation either being separate from the entity or comprising just some of the entity's attributes.

One object-based design of a GIS is the Digital Line Graph-Enhanced (DLG-E) feature-based model, a new modeling approach of the United States Geological Survey (USGS) (Guptill 1990). The USGS realized the Digital Line Graph (DLG) format (USGS 1989) produced for cartographic data was no longer meeting the demands placed on it (Guptill 1988). This motivated building a feature layer on top of the DLG-5 topological structure of nodes, lines and areas.

The DLG-E enhances the previous DLG model in many ways. One improvement is the ability to reference particular features by name. For example, a user can refer to a stream as Broad Brook Stream. Broad Brook Stream is an example of a "feature instance." A feature instance is composed of a "feature object" having non-spatial attributes and (if it is to be mapped) one or more "spatial objects" having locational data. Another enhancement in DLG-E is the definition of relationship types between objects. Relationship types have been predefined between an object and its spatial object(s), between spatial objects, and between feature objects. Example DLG-E relationships between feature objects include: composed_of, outflow_to and connected_to.

Designing a GIS by focusing on individual entities (e.g., a stream, a bridge or a well) rather than focusing on the geometric representation of a map as is now done (e.g., storing raster cells or vector components) will have a number of ramifications. Specifically, the feature-based model promotes:

- An increased amount of non-spatial information to be stored within the feature object,
- More than one spatial representation to be associated with each feature object (e.g., different scales), and
- Relationships between feature objects to be directly stored as part of the model.

We anticipate more non-spatial data will be stored within each feature object because of the new focus on features. As to multiple geometries, (Newell 1992) has also noted this as a benefit of object-centered models. In addition, the DLG-E feature-based model, in particular, designates and stores relationships between objects. The increase in information will result in more comprehensive and useful databases, and an OODBMS implementation promotes this advantage.

**An OODBMS Design for a Feature-Based GIS**

In this section an object-oriented DBMS is used as a framework for modeling a feature-based GIS. The modeling of feature objects, relationships, and spatial objects is shown using classes with encapsulated member functions and having user-defined types, reference pointers, and sets as data members.

**Feature Objects.**

Each DLG-E feature object is modeled as an object in an object-oriented system. For example, Broad Brook and Pheasant Streams are individual objects of the class of
stream objects (Figure 1). As part of the DLG-E design (Guptill 1990), attributes have already been identified and listed for each mapped feature object. These attributes are the data members of a class definition for that type of object. For instance, Broad Brook Stream is intermittent. Methods for each class are also included in the class definition. For example, a class containing objects with a spatial representation includes a DISPLAY function.

To implement these objects in an object-oriented database system, each object is declared to be persistent. Furthermore, objects in a class may be physically and/or logically clustered for efficient access to objects of that type. In Figure 2, Broad Brook and Pheasant Streams are clustered along with other stream objects in the class of streams. Similarly, all bridges are clustered, and all line objects are clustered.

Relationships between Objects.

DLG-E has designated possible relationships between feature objects and spatial objects (feature-spatial), between feature objects and other feature objects (feature-feature), and between spatial objects and other spatial objects (spatial-spatial). In an OODBMS, reference pointers (defined earlier) are used to model these relationships.

The DLG-E feature-spatial relationship between a feature object and one or more spatial objects is the "composed_of/part_of" relationship. In Figure 1, the multiple-scaled feature object Broad Brook Stream has two spatial representations, one as a line and one as an area. In Figure 2, the feature-spatial relationship is shown again.

Modeling the DLG-E feature-feature object relationships also uses reference pointers. For example, in Figure 1 Broad Brook Stream "outflows_to" Pheasant Stream. Not shown is the DLG-E aggregating feature-feature relationship in which a compound object of type watercourse is "composed_of" basic stream objects, including Broad Brook Stream. The relationship shown in Figure 2 in which Tomahawk Bridge "spans" Broad Brook Stream is not one of the five predefined DLG-E feature relationships. However, the use of reference pointers in an object-oriented model easily allows any type and number of relationships between feature objects. Because of this, an expanded view of modeling relationships in the feature-based model is taken in this paper.

Which additional relationships, other than those predefined in the DLG-E model, to explicitly establish between feature objects and whether they should be unidirectional or bidirectional must be decided by the application developer for the specific use of the database. For example, additional feature-feature relationships can add significant meaning to the database and/or facilitate query processing. Once decided, some of those relationships might be determined by running spatial algorithms (e.g., to find borings drilled alongside roads). Others may be determined by matching values between feature classes using, for example, ODE's join facility (e.g., to find what highways are in which counties by matching on county name). Relationships needed but not able to be set by algorithms will have to be done by the database developer. Design decisions to minimize redundancy, maximize efficiency, and establish relationships need to be made with respect to the anticipated queries posed to the database.

FIGURE 1. Object-oriented design for DLG-E feature objects, spatial objects, and relationships
Reference pointers also allow relationships for a hierarchical categorization or aggregation of feature objects. Currently, in DLG-E only one level of aggregation is permitted. That is, the compound objects watercourse and route are composed of lake/pond and stream/river objects or road objects, respectively. However, this can be expanded to allow any number of levels of aggregation using OODBMS pointers. The ease of modeling groups and subgroups provides more meaning to be added to the data. In Figure 3, a categorization hierarchy is shown above feature objects using the DLG-E grouping categories called views. Meta-objects are created for each view level. A building, a bridge and a well are examples of feature objects that are structures in the Built-up Land subview of the Cover view.

Spatial Objects and Relationships.
Our model designates separate spatial objects rather than embedding spatial information within a feature object. This means a feature object does not inherit from a general spatial object of its type as, for example, in (Milne 1993). Although storing non-spatial and spatial data within a feature object would be efficient for queries that address both kinds of data, it is a less efficient organization for retrieving either kind of data separately because of the increased quantity of information within each object. Because fast access is especially needed for spatial data to perform spatial algorithms and to do mapping, we have kept spatial information separate. In our design, spatial objects inherit attributes and methods from a general spatial object of that type.

DLG-E allows for the four spatial objects of point (degree=0), node (degree ≥1), chain, and polygon. In general, other spatial representations are possible including a raster (grid-based) representation or variations on topological structures with points, lines and polygons.

DLG-E defines specific relationships between spatial objects. For instance, a point is “within” a polygon, a node “bounds” a chain, and a chain “bounds” a polygon. Relationships such as these are modeled using reference pointers in an OODBMS.

Changing a GIS Application From a Relational DBMS into a Feature-Based OODBMS

This section uses the GeoGIS data set to illustrate a feature-based model in an OODBMS and to show modeling advantages of an OODBMS compared to a relational DBMS. The database considered here exemplifies a GIS with subsurface information associated with point locations in a geodetic coordinate system. Using the point data type illustrates issues associated with implementing the feature-based model. First, a subset of the database is presented to illustrate the approach and how an OODBMS eliminates relational tables. Then, the entire GeoGIS model is shown in an OODBMS version.

Relational DBMS Model of GeoGIS

In GeoGIS, geographic data for base-map coverages of highways, townships, and hydrography in Dane County, Wisconsin is integrated with data from drilling
and sampling subsurface strata (Adams 1993). Figure 4 contains only a subset of the data in the 16 relational tables (shown in Figure 5) implemented in the INGRES relational database management system (Stonebraker 1976).

The Entity Relationship (ER) Diagram shows the basic entities and relationships before a specific implementation model is chosen (Figure 4a). Each boring is composed of a set of segments (the soil segments found in drilling the well) and a location. For the relational model, all boring logs are listed in the Boring relation (Figure 4b). Since the relational model cannot directly handle a set of segments being embedded within a boring tuple, segments must be placed in another table. The Segment table, then, has boring# as a foreign key to relate back to the Boring table. In the GeoGIS database, location information is kept in the separate relation BoringLoc. This was necessary because spatial queries were done on the data using the GIS GRASS (Westervelt 1991), which has a limited interface with relational databases.

Because segments and borings are in separate tables, the tables must be related back together for queries. For example, to find the soil description of the top segment for borings drilled by the DNR, Boring and Segment are joined on boring#. Using SQL (Structured Query Language), the query would be:

//SQL query/
SELECT agency, boring#, description
FROM Boring, Segment
WHERE Boring.agency = "DNR" and
  Boring.boring# = Segment.boring# and
  Segment.segment# = 1

Although the relational model supports high level relational algebra and calculus languages, if many tables need to be joined, queries are cumbersome to write, and data must be re-assembled from storage.

**OODBMS Model of GeoGIS**

Using the OODBMS set construct, segments can be embedded within borings, and the object-oriented schema of GeoGIS contains just two feature classes. Borings form one feature class (boringclass), and point location data is the other (boringloclass). Using ER notation in which boxes indicate entity types (classes) and ovals denote attributes (data members), Figure 4c shows a model for one boring object and its spatial representation object. The set of segments is shown in an oval just as any other attribute of a boring. A segment is a user-defined type consisting only of segment# and description because the foreign key, boring#, is no longer needed. Also, now, the feature-spatial relationship, composed_of, can be directly modeled with a reference.
FIGURE 4. GeoGIS Schema a) ER diagram (boxes indicate entities, ovals-attributes, diamonds-relationships) b) Relational model c) Object-oriented model (drawn for one boring object and its spatial object)

(a)

Boring (subset)

<table>
<thead>
<tr>
<th>boring#</th>
<th>agency</th>
<th>date</th>
</tr>
</thead>
</table>

Segment

<table>
<thead>
<tr>
<th>segment#</th>
<th>boring#</th>
<th>description</th>
</tr>
</thead>
</table>

BoringLoc

<table>
<thead>
<tr>
<th>boring#</th>
<th>x</th>
<th>y</th>
<th>elevation</th>
</tr>
</thead>
</table>

(b)

(c)

location object

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>elevation</th>
</tr>
</thead>
</table>

boring object

<table>
<thead>
<tr>
<th>boring#</th>
<th>agency</th>
<th>date</th>
</tr>
</thead>
</table>

spatial ptr

(set of embedded segments)

<table>
<thead>
<tr>
<th>segment#</th>
<th>description</th>
</tr>
</thead>
</table>
pointer from the boring object to its spatial object. The schema for boringclass has the data members:

boring#, agency, date, reference pointer to a spatial point object, and a set of segments, segmentset.

Comparison of Relational and Object-Oriented DBMS Modeling-Eliminating Tables

Before presenting the entire GeoGIS database, this section discusses modeling capabilities of OODBMSs that eliminate extra tables needed for standard relational DBMSs. As used here, extra, or secondary, tables are those no: corresponding to a main entity or feature type. An analogy to a relational table is a class in the object-oriented model. Fewer feature classes will result in the object-oriented version. This is because, in an object-oriented DBMS, a data member of an object can be:

- a complex user-defined type,
- a set of any type, or
- a reference pointer to indicate a relationship between objects, e.g., to point to a subobject for the composed of relationship.

Having fewer tables results in a simpler schema and a focus on the main feature types.

Being able to define complex user-defined types instead of being restricted to simple types such as integer or character in the basic relational model does not directly eliminate extra tables. However, without being able to group attributes, the database designer may elect to create an additional table just to keep logically related information together. For example, without a user-defined type it may be a better logical design to place the spatial attributes x, y, and elevation in the separate spatial relation BoringLoc even if it is not needed for query efficiency. Having user-defined types is so useful, it is one of the first features added to extended relational DBMSs.

Being able to have an attribute be a set eliminates a table because, in the relational model, an attribute can only be single-valued. As a result, for a multi-valued attribute, normalization requires creating a separate table that has one tuple for each attribute value along with a foreign key to relate back to the original table. Thus, an extra table results for every multi-valued attribute. The Segment table in the relational GeoGIS database is an example of sets of values. Contrary to this, OODBMS set constructs allow multiple values to be modeled within the entity they are describing, resulting in a more natural representation. In actual implementations, currently most OODBMSs only allow pointers to be set elements, but the intended set values can be stored nearby and the logical model is preserved.

In addition, an object-oriented DBMS eliminates the extra tables needed to model relationships. In the relational model, relationships between objects are modeled by associating the key values of each entity in a separate relationship table. Instead, object-oriented systems are "reference" rather than value-oriented; that is, objects are accessed using OIDs and not attribute values (Cheng 1991). For example, as shown in Figure 4c, location data is now directly accessible from a boring using a reference pointer rather than a foreign key. For some queries, then, retrieval time is minimal compared to the relational model in which searching for key values is needed.

Furthermore, even more tables are eliminated because the potentially sophisticated process of normalization into higher forms (e.g., 2NF, 3NF, BCNF, etc.) to avoid update design anomalies within tables such as partial, transitive, and other dependencies is not needed.

The Entire GeoGIS Database

Figure 5 shows the complete set of tables for the GeoGIS relational database along with the object classes for the OODBMS translation. For simplicity in showing the attributes in the relational version, usually only the foreign key is listed. The object-oriented version results in just three feature classes that correspond to the main entity relational tables for borings, samples, and boring locations. All of the remaining relational tables contain further information about the main entity types and can now be eliminated. For example, in the object-oriented version, the results of all the sample tests are values in user-defined types embedded within a sample. Depth information for segments and samples is now stored within those objects to be appropriately near the re-arranged data. The sampler type tables are now within a sample as part of a programming language structure with variant elements (a union).

In GeoGIS, each boring has a set of soil samples taken from it. Because independent analyses are run on samples, it is not efficient to embed them within a boring. Instead, samples are separate feature objects related to borings with a feature-feature relationship. With reference pointers and sets available in an OODBMS, this is modeled with each boring containing a set of pointers to its samples.

As can be seen, the use of OODBMS user-defined types, sets, and reference pointers has advantages over
FIGURE 5. Translation of GeoGIS from a Relational DBMS to an Object-Oriented DBMS

(a) Complete GeoGIS relational database (* denotes subset of attributes) b) GeoGIS in an OODBMS

Boring (boring#, agency, date)*
BoringLoc (boring#, x, y, elevation)
Sample (sample#, boring#, sampler-type)*
SampleLoc (sample#, depth)*
Segment (segment#, boring#, description)*
SegmentLoc (segment#, d_bottom)*

Sampler Types
SPT (sample#)*
Shelby (sample#)*
Other (sample#)*

Sample Tests
Atterberg Limits (sample#)*
Gradation (sample#)*
WaterContent (sample#)*
Strength (sample#)*
UnitWeight (sample#)*
SpecificGravity (sample#)*
Consolidation (sample#)*

Cluster of Borings
boring#
agency
date
pointer to boring location
{ set of segments }
{ set of pointers to samples }

Cluster of samples
sample#
sampler_type
dept
AtterbergLimits test
Gradation test
etc.

Cluster of Boring Locations
x y elevation

30 URISA Journal/Refereed
the relational model. First, tables are eliminated making the schema simpler to understand. The meaning of the data and relationships was not apparent from the GeoGIS relational tables, whereas the organization in the OODBMS more directly represents the way the data is actually related. Query writing will access fewer tables. In addition, updates can be easier. Although usually a key value does not change, if a sampleID changes in the object-oriented version of the database, only one attribute in one place needs updating. However, in the relational model, because sampleID is a foreign key, 11 additional tables would need to be searched for that sampleID to change it. In general, the use of pointers expedites updates between related objects.

Compared to the relational version, the feature-based OODBMS model of GeoGIS has fewer classes (tables), giving a focus on the main feature types. Relationships between classes are clear as they are explicitly stored.

Other Object-Oriented DBMS Features for a GIS

Besides the object-oriented modeling allowed by the underlying programming language and DBMS extensions, there are other useful DBMS features for a GIS. Basic to this is having data be persistent. The syntax of ODE's language, O++, is used throughout this section to illustrate programming the GeoGIS application in an OODBMS.

Persistence

Because the database language O++ developed from the non-persistent object-oriented programming language C++, to store GeoGIS objects in the database, they must be specifically declared as persistent using an extension of O++. Persistence is the ability of an object to survive from one run of a program to the next run. By declaring an object to be persistent, the programmer does not have to read or write to external files. Persistence is essential for a database system (Richardson 1989).

O++ achieves the goal of persistence being orthogonal to types; that is, an individual object is persistent, not an entire class. Persistent objects are allocated using pointers and the function pnew. For example, to make a boring object be persistent, it would be allocated this way in O++:

```c
persistent boringclass *p;
p = pnew boringclass(100, "DNR",
    "July 8, 1990", etc.);
```

Clusters of Objects

Just as all tuples in a relation are usually stored together, an OODBMS also provides for the physical clustering of objects. ODE automatically groups all persistent objects of a class into clusters, and other OODBMS have provisions allowing the user to cluster related data, sometimes of different classes, for efficient retrieval. The clustering of objects in a class is beneficial for queries with an access pattern to retrieve similar objects, e.g., map all well points or find all bridges with a span greater than 50 feet.

In ODE, executing a create function on a type name insures all objects of that type will be stored together. For instance, before the first allocation of a boring object (e.g., using pnew as above), a create statement needs to be executed:

```c
/*create a cluster for boring objects*/
int create (boringclass);

In ODE, subclusters were designed to improve access efficiency. That is, within a cluster, subsets of objects can be physically grouped. For example, within the cluster for boringclass, borings can be grouped by agency such as the Department of Natural Resources (DNR):

```c
/*create a DNR subcluster*/
create (boringclass::"DNR");
/*store a new boring within its subcluster*/
p = pnew boringclass (100, "DNR",
    "July 8, 1990", etc)::DNR;
```

Querying Data

OODBMS provide set-level querying facilitates similar to relational query languages. Thus, access to objects is at a higher level than by individual object ID (OID). To access objects in sets or clusters, ODE provides an iterator which is a special form of the "for" loop. The iterator also includes "suchthat" and "by" clauses to be passed to the object manager to select only the OID wanted, thus allowing optimization.

Using nested iterators, here is the ODE version of the SQL query given earlier to print the soil description of the top segment for borings drilled by the DNR.

```c
/* O++ iterators */
for (b in boringclass)
suchthat (strcmp(b->agency, "DNR") == 0)
    for (s in b->segmentset)
suchthat (s->segment# == 1)
printf("%s %d %s\n", b->agency,
b->boring#, s->description);
```

The above query is not as high level as SQL, but a standard high-level query language does not yet exist.
for object-oriented systems (Elmasri 1989). However, in the proposed SQL-like interface to ODE, CQL++ (Dar 1992), the above query would be:

```c
t community ODE's CQL++ */
SELECT agency, boring#,
   segmentset.description
FROM Boring
WHERE agency = "DNR" and
   segmentset.segment# = 1
```

Although dot notation is needed for a complex attribute, this query is simpler than its SQL equivalent for the relational model because only one table is referenced. Also, executing the query will have potentially fewer disk accesses because when a boring drilled by the DNR is found, its segments are included or are nearby. Contrary to this, in the relational model, boarings of DNR boarings have to be matched with tuples in the Segment table which will involve more disk accesses even if an index is used.

In ODE, individual persistent objects can be named. Named persistent objects can be accessed quickly as iteration is not done.

As in the relational model, but not standard in OODBMS, ODE provides for arbitrary joins between classes. That is, an iterator can range over multiple clusters. For example, to find samples taken at the greatest depth of a boring, one would write: (Note in Figure 5, samples form a separate class in the database. Also, the terminal depth is an attribute of a boring that is not shown.)

```c
/* An ODE join */
for (b in boringclass; s in sampleclass)
suchthat (b->term_depth == s->depth)
   printf ("%s %d %s
", b->boring#, s->sample#, s->depth);
```

### Inheritance and Virtual Functions

Inheritance is provided by the object-oriented programming language of the OODBMS but is included in this section along with other illustrations of code. Inheritance is useful for GIS non-spatial and spatial data. The following example shows inheritance for spatial data. Again, our design of the feature-based model has spatial information forming separate objects from the feature object. Each spatial object inherits from a generic spatial object of that type. For example, a generic point class exists from which specific point data, e.g., for a boring, inherits attributes and methods. The declaration in C++ of a point base class is:

```c
class pointclass { //base class
public:
   float x; //x geodetic coordinate
   float y; //y geodetic coordinate

   /* constructor */
   pointclass(float x, float y);
   /* default display routine */
   virtual void display();
   /* standard buffer routine */
   void draw_buffer(float);
};
```

The spatial object of a boring inherits from pointclass as follows:

```c
/* derived class */
class boringloclass : public pointclass {
public:
   float elevation; //surface elevation
   /*constructor*/
   boringloclass(float x, float y,
                 float elevation);
   /* code to display wells in a special way */
   void display();
};
```

Boringloclass is a derived class of pointclass and inherits the data members, x and y, and the member function, draw_buffer. It added elevation to complete its location data. The constructor includes the inherited x and y data members. Since the display function is virtual in pointclass, it would have only been inherited if the derived class had not supplied its own version. In this way, C++ allows a form of polymorphism. But here, boringloclass did define its own special version of display. Boringloclass does inherit draw_buffer from pointclass.

### Constraints

Constraints are boolean expressions used to maintain consistency in the database. They are useful in a GIS database for error checking. For example, a constraint to ensure the agency name of a boring is either "DNR" or "DOT" would be:

```c
constraint:
   agency == "DNR" || agency == "DOT";
```

Constraints are specified in the class definition and are checked at the end of constructor and member function calls on an object. If a violation of a constraint occurs, the transaction (the set of current statements executed by the user) is aborted and rolled back. Constraints are inherited and can be used to specialize classes. Constraints will be in future releases of O++. 
Updates and a “Smart” Database

Triggers are actions that are executed when a condition becomes true in the database. They greatly help in update situations because they execute methods that perform possibly cascading changes among objects. Triggers, available in future releases of O+++, are listed in a class definition and are either once-only or perpetual.

Triggers produce a “smart” database by automatically making a series of changes. For example, an O++ trigger called “changed” will detect a change in an attribute. With the GeoGIS database, if the trigger detects “changed(date)” to be true in an update transaction on a boring, a function will be called to automatically make other necessary date changes in the database. As another example, locational changes are needed for cadastral objects when the location of a reference object is updated (Kjern 1990). This is a situation in which a full-fledged DBMS with triggers would be quite useful.

Although triggers exist in the relational model, an object-oriented design facilitates the work needed to be done by the trigger because reference pointers directly locate objects that are affected rather than having to search tables for key values.

Versioning

ODE provides for versions of an object. As with persistence, a version is an object property, not a class property. Being able to save multiple copies of an object is important to keep historical values as well as unique versions of an object. This is important for GIS temporal data and data lineage.

Summary

The Advantages of Object-Oriented Modeling

In this paper, we have shown that object-oriented modeling aptly suits the new feature-based GIS model. A feature object is an instance of a class for its type of entity set. The various non-spatial attributes are data members of the class, and specific methods for the class are written. Each mapped feature object has one or more reference pointers to its spatial object representations. Other pointers show relationships between feature objects (e.g., outflow_to or connected_to). Pointers that express the part of relationship allow aggregation of sub-objects. The ease of modeling relationships using reference pointers expands the feature-based model’s focus on explicitly stored relationships. Specific types of spatial objects representing locations of features inherit spatial attributes and methods from generic spatial classes.

An OODBMS has advantages over a standard relational DBMS. Use of an object-oriented programming language in the OODBMS provides inheritance, user-defined types, and the pointer data type. Inheritance simplifies class definitions by avoiding repetition of common fields and methods for specialized types. User-defined types allow for complex attributes which logically group information. Pointers promote the storage of relationships between entities and can facilitate updates. In addition, OODBMSs provide a set construct allowing an attribute to consist of many values.

These advantages eliminate extra tables needed for a relational DBMS. As used here, extra tables are those not directly corresponding to a main entity type, e.g., relationship tables, tables formed to eliminate multi-valued attributes, and tables produced to eliminate update anomalies. A relational table is roughly analogous to a class of objects. Having fewer tables (i.e., classes) has a number of performance and logical advantages. Fewer disk accesses are needed to retrieve all information if values are embedded in the main object. Also, direct references to related objects using pointers can eliminate a number of the disk accesses involved in searching and joining tables for some types of queries. Harris describes this performance advantage of OODBMS over relational DBMS (Brodie 1990). Work on pointer-based join methods and their performance is reported in (Shekita 1990).

In addition, logically, the overall schema of the database is easier for the user to understand and remember without tables that do not correspond to main entity types. Queries are easier to write as the user does not have to keep track of many tables and how they are related and will not have to write a sequence of joins between tables. The relational model becomes quite cumbersome with many tables related by foreign keys.

Also, we expect the impact of a feature-based GIS model to be that great quantities of additional information will be stored in the database. Besides more non-spatial attributes, more relationships between feature objects will be stored. Having additional data will exacerbate the problems of extra tables needed for the relational model.

It can be speculated that spatial queries and mapping may be faster with pointer-following algorithms rather than value-based joins. For example, using a relational system to model a line and its corresponding line segments could result in a separate tuple for every line segment, e.g., (Haas 1991). Then, finding the individual line segments can only be done by searching on a foreign key value. Indexes would speed up the search time. But, retrieving line segments in an OODBMS could be much faster for some queries because line segments are directly referenced using pointers.
Compared to a traditional GIS, using an OODBMS for a GIS allows spatial and non-spatial data to be in the same database. In addition, the concept of layers is looser because all feature objects (and spatial objects) are together. Because of this, it will be easier to retrieve information. Also, complex sharing of spatial objects can be done, as also noted by (Newell 1992).

**Summary of DBMS Features of ODE that Facilitate a GIS**

In ODE, all data can be persistent. Note, this is contrary to a system as in (Kjerne 1990) built using Smalltalk-80 which is object-oriented but without the persistence of a DBMS. In ODE, persistence is not a property of class; each object must be allocated as persistent. Thus, the modeling of persistence as suggested in (Egenhofer 1992) would not be done by inheritance from a database superclass. Being able to allocate objects of any type in either persistent or volatile store will be useful for GIS decision-making involving various potential features or their locations.

Clusters group objects of a class and are important for efficient access. Subclusters offer additional efficiency by allowing the user to store subgroups of objects that have the same value for some field. Iterators approach a high level query-type access for clusters of objects. Named persistent objects allow fast access to particular objects, and joins can be done using the associative for loop.

Integrity constraints provide automatic error checking by the DBMS on the user's data to insure consistency. Triggers are a useful DBMS feature to automatically propagate changes in update situations when a change affects more than one value or object. Versions allow multiple copies of an object. Transaction management and hypothetical transactions have now been added to ODE.

The above features provide an adequate software framework for implementing a design for a feature-based GIS. Vendors are still working to incorporate the full set of standard DBMS features into OODBMSs which eventually will be beneficial for a complete, multi-user GIS.

**Extensibility of ODE**

Previously, it was noted that current DBMSs need to provide extensibility besides better modeling features to meet the needs of a GIS. The following discussion compares the table presented earlier to what is provided by ODE for building a feature-based GIS.

Extensibility for adding new data types and operations is provided by the O+/C++ class and methods. Abstract data types (type definitions plus operators) can be defined for DLG-E feature objects and spatial objects. For instance, a point class can be declared that includes operations such as shortest distance between points. Other spatial operations such as overlay can be written as methods of a class or as general functions. Regarding access to user-defined code and other packages, many routines can be added or accessed because the programmer has C++ with which to work.

As to the other needs and overall extensibility for a spatial system such as a GIS, ODE is a general purpose OODBMS; it was not designed specifically for spatial data. However, because access to ODE is a programmer's interface, most of the following needs could be designed and programmed by the GIS implementor. GISs need spatial objects (e.g., points, lines, and areas), spatial operators, spatial storage and access methods, spatial indexes, and appropriate optimization.

The parts of ODE that are not inherently extensible are storage methods and optimization. Currently, ODE clusters all objects of the same class but more sophisticated clustering or spatial storage structures might be needed. Also, regarding optimization, the designers of ODE have not yet addressed extensible optimizer issues for new operators and indexes.

Thus, the programmer/designer of a GIS using ODE could declare feature objects and spatial objects, write spatial methods, establish relationships between feature and/or spatial objects, code spatial indexes, and write queries in O++.

Eventually, an OODBMS could provide built-in features for a spatial system (spatial indexes, spatial storage structures, spatial query operators and optimization), but for now, an extensible object-oriented DBMS forms a good framework for prototyping a feature-based GIS.

**Acknowledgments**

This material is based upon work supported by the National Science Foundation under Award No. MSM-9010587. Dr. Ken Chong, cognizant program director. The authors wish to thank Dr. Laura Haas of IBM Almaden Research Center for her comments and suggestions on early drafts and AT&T Bell Laboratories for their contribution of the ODE Object Database and Environment (Agrawal 1991). This work was partially supported by an American Fellowship to the first author from the American Association of University Women.

**References**


The National Spatial Data Infrastructure: Designing Navigational Strategies

Steven Frank

Abstract: The place of digital geographic data and land information in the evolving electronic information infrastructure is of great concern to many in the GIS/LIS field. The way we structure spatial data and information (spatial resources) within the architecture of the information infrastructure may determine both the manners in which those resources may be used and the design of parts of the infrastructure architecture itself. A part of this architecture should include navigational tools that may be easily used to efficiently and effectively locate and access the information contained in the infrastructure.

Cataloging has long been recognized as an efficient method for condensing knowledge of large collections of items. Special approaches to cataloging have proven successful for many collections, including collections of spatial resources. The advent of interconnecting computers and new network information retrieval tools calls for an examination of catalog-like services that these new tools might provide. In this paper, we examine several likely scenarios for implementation of the information infrastructure as it pertains to spatial resources. We then develop cataloging paradigms to examine how the needs of spatial resource searchers may be met in possible infrastructure scenarios using the concepts provided by these new networking retrieval tools.

The ongoing convergence of computer networks with public telephone networks and cable television distribution systems is expected to bring about a national, public-broadband information infrastructure available to all (Kapor and Berman 1992). The place of geographic and land information in the information infrastructure has become a topic of great interest in recent months. The ability to rapidly exchange spatial information and data across the infrastructure is seen as a key element to global harmony, health and well-being (NRC 1993). At the national level, the Federal Geographic Data Committee (FGDC) has been working with some success toward bringing together federal, local and private interests to form a National Spatial Data Infrastructure (NSDI). One of the key issues for development of NSDI will be the ability of users to rapidly find spatial data and information of interest (NRC 1993; FGDC 1991). Since NSDI will almost certainly be but a part of the information infrastructure, users will also require mechanisms that will allow them to both differentiate and integrate spatial data and information, or spatial resources, from the myriad other information sources available over the infrastructure.

Spatial resources can exist in a number of forms. A spatial resource might consist of files or databases containing digital spatial data, or it might consist of tools such as software to manipulate digital spatial data. Alternatively, a spatial resource might contain spatial data products derived from spatial data, such as maps and tables. Finally, a spatial resource might perform spatial data services, such as automated navigation updating for vehicle location. A vital spatial data service would be spatial metadata catalogs to locate, describe, and access spatial resources.

We concentrate on spatial metadata catalogs in this paper, but note that in the evolving electronic information infrastructure, the distinction between how we view catalogs and other resources is becoming less clear than in our paper-based society. The ability to locate, access and use on-line spatial information in a single computer session points to a need to develop tools that can smoothly and efficiently intermesh these functions. Thus a user should eventually be able to locate a relevant spatial resource, access that resource, and use it in a productive manner while being only marginally aware of the diverse operations involved.

In the first section of this paper, we discuss how users locate information. Next, we describe some of the information infrastructure tools currently being used to locate information. We then discuss the difficulties in infrastructure "navigation." The following sections outline various scenarios for designing a spatial data infrastructure and how such scenarios may affect the design of useful navigation tools.

Steven Frank is a PhD candidate in the Department of Surveying Engineering at the University of Maine, Orono, and is a research assistant with the National Center for Geographic Information and Analysis.
Locating Information

The manner in which users find information can be categorized by six methods:

1) Retrieval—the user knows what information is needed, knows where that information resides, and understands how to obtain it;
2) Searching—the user knows what information is needed, believes that the information is in the system, and is unsure how to obtain it;
3) Browsing—the user is unsure what information is needed and is unsure if the information is in the system;
4) Scanning—the user is not looking for particular information but is scanning for interesting items;
5) Exploring—the user is testing the limits, constraints, and capabilities of the system; and
6) Wandering—the user has no structure in his or her search and does not understand the system which is being accessed (Stephenson 1988).

Users need the flexibility to easily find and access information in a manner best suited for a particular situation. For example, a geographic data user may wish to retrieve a certain spatial dataset, then search or browse for related datasets.

Numerous spatial data cataloging systems and spatial database systems have been built to date. However, only a few of these systems are attempting to interoperate with one another or with other spatial resources, such as on-line software applications (Schneider and Murray 1993). None of these systems seems to offer the full range of methods needed by users to effectively locate spatial resources (Frank 1994). Despite these developments, many people searching for geographic data still rely on informal contacts with other geographic data users using methods such as electronic bulletin boards and discussion groups—such as the Geographic Information Systems Discussion List (GIS-L) and the Maps and Air Photo Systems Forum (MAPS-L).

Some spatial-data cataloging systems are experimenting with recently developed general information infrastructure retrieval tools to determine the suitability of those tools for locating and accessing spatial resources. Brief descriptions of several current network tools and some of the spatial data cataloging systems that utilize those tools are given below.

Internet Tools

Many tools for locating and retrieving digital information have been developed or are currently under development. Some of these tools, such as File Transfer Protocols and TELNET, have become de facto standards for interchanging information between computers on the Internet. The Internet Engineering Task Force (IETF), which oversees the development of Internet standards and protocols, expects to have standards in place for Universal Resource Numbers (URNs), which give all on-line resources a unique identity, and Universal Resource Locators (URLs), which point to all instances of URNs, in place within the near future. IETF is also working on protocols to make many of these existing tools interoperable (IETF 1993).

Anonymous FTP

One method of assessing on-line information is the Internet File Transfer Protocol (FTP). FTP software allows users to send and retrieve files of any type, but provides no methods to search for those files (Kahn 1992). Special public FTP sites are available to anyone with access to the Internet through a process known as “anonymous” FTP. Numerous anonymous FTP sites offer digital spatial data and software for processing spatial data (Nyman and Sealy 1993), including the National Center for Geographic Information and Analysis at Santa Barbara, California. Users must serially access FTP sites to retrieve files. Browsing or scanning for possibly useful files is also done by serially accessing sites.

TELNET

Another popular Internet application is TELNET. TELNET software allows users to connect to databases from remote locations (Kahn 1992). TELNET does not allow connection to more than one database at a time, therefore users must again jump from site to site in search of information. Most TELNET sessions allow users to log in to remote terminals where databases services, such as spatial data catalogs, are available. Interfaces to these databases normally vary and users must often learn several interfaces to access different databases. Users must again access TELNET sites serially, although some systems, particularly those in the NASA Earth Observation System Data and Information Systems (EOSDIS), allow users to interactively move from catalog to catalog through the interface menus. The Global Change Master Directory lists the earth science datasets held in EOS Data Active Archive Centers (DAACs) (Strand 1992; McDonald and Blake 1991; Ramapriyan and McConnaghy 1991); however the system is limited to earth science data.

TELNET and related remote log-in services can also provide on-line access to spatial databases themselves. One may operate a GIS or LIS remotely from a terminal almost as if the system was residing on that terminal.

More sophisticated networking tools based on TELNET protocols—such as Gopher, Wide Area Information Servers, and World-Wide Web—have been developed and are currently popular with many Internet users.
Archie

An automated service, called “archie,” has been developed to allow users to search for files accessible by anonymous FTP. Archie is available at several TELNET sites. Users are required to know all or part of a file name in order to perform a search. Archie proactively retrieves and indexes files stored in anonymous FTP sites. Users can then query archie for the existence of files matching filename criteria. Archie will return a result set of electronic address of files meeting the criteria (Duetsch 1992). Duplicate files existing at different sites will be duplicated in the result set. Archie also does not distinguish between files and directories, so the result set may also contain the addresses of directories whose names match the given criteria. Once a site containing a specific instance of a file is located, it may be browsed for other relevant information, but there is no way to ascertain that all sites containing relevant information have been located if a file name is not known or if references to the file name cannot be found.

Gopher

Gopher systems organize on-line data and information as a hierarchy of directories and files useful for browsing or scanning information. Search mechanisms for gopher, called “veronica” and “jughead,” allow users to search not only for files, but also for services provided by various Gopher implementation (Wiggins 1993; Alberti et al. 1991). Gopher was developed primarily to provide information on an organization-by-organization basis. A new gopher development, called “gopher+,” allows servers to attach added information to gopher resources, such as file format and publishers’ name, to enhance searches for information. Gopher+ is designed to be downwarly compatible with the non-enhanced versions of the software. Gopher was developed using a client-server model. Client interfaces vary widely in appearance as do the presentation of the information itself. Clients may use internal algorithms to display information or they may launch external processes, such as image browsers or audio software to play music. However, all Gopher clients receive the same information (Wiggins 1993).

Gopher can be used to access data catalog information at the Consortium for International Earth Science Information Network (CIESIN). Gopher gateways allow TELNET access to several spatial data cataloging systems from within gopher.

Wide Area Information Servers

The Wide Area Information Server (WAIS) is a full-text retrieval system based on keyword searches. Users may simultaneously search multiple sites and may expand or narrow their searches using the concept of “relevance feedback,” i.e., using the results of previous queries to formulate new queries. The Wide Area Information Server is a “stateless” system, meaning that every transaction between nodes is separate and complete from every other transaction. The primary advantage of stateless systems is that they greatly reduce the amount of network traffic that can be incurred when using other network retrieval tools (Machovec 1992; Kahle and Medlar 1992). The U.S. Geological Survey (USGS) has implemented a data catalog, called the Earth Science Data Directory (ESDD), using WAIS that is only search-able using keyword, not spatial indexing, constraints. USGS is reportedly working on a WAIS model that will allow retrieval based on spatial restraints.

World Wide Web

The World-Wide Web (WWW or W3) retrieves on-line information through the use of hyperlinks. The links are made using marked-up text to tag documents and objects within documents using a Standard General Markup Language (SGML). These tags can contain pointers to external resources that are logically connected to the tagged item. The user can first access an index of documents that may be searched and then link to a relevant document listed in the index. As the user scans the selected document, new hyperlinks will appear for related documents which may also be retrieved and browsed. WWW is also a stateless system, but does not currently support relevance feedback (Berners-Lee et al. 1992).

X.500 Directory Service

The X.500 Directory Service serves as a “white pages” for network users. The X.500 protocol is intended to be a “standard for a global, logically centralized but physically distributed, electronic network directory” (Planka 1990, p. 94). The X.500 Directory Service is intended primarily to locate and index people and resources much like a telephone directory rather than to provide the more sophisticated full-text searching and browsing described above.

Navigating the Infrastructure

The ability to find and access on-line resources in the National Information Infrastructure is often termed “infrastructure navigation” (Kahin 1992; Lynch and Preston 1992). Successful navigating strategies should offer users both flexibility and structure in performing their searches for resources, some uniformity of access to resources, and a clear understanding of the possible or permissible uses of a particular resource. The NII currently offers few useful tools based on spatial resource characteristics for navigating the infrastructure. While

Frank/LIRISA Journal 39
some spatial data is available using network services, it is still difficult to find. It is also difficult to understand what one has found when a spatial resource is accessed without a priori knowledge of what that resource contains and how it works.

The problem of finding and accessing relevant spatial information is not an indexing or cataloging problem alone. Numerous on-line spatial data catalogs, data collections and land information databases already exist. Formal methods for interrelating these systems are needed to allow a framework for building the organizational, economic and legal structures that are necessary to gain a common acceptance and use of the infrastructure. These methods, in turn, will depend on the development and nature of the infrastructure. To develop an understanding of what these methods might be, it is advantageous to examine various possible development scenarios of NSDI to analyze how such methods might affect successful implementation.

Spatial Data Infrastructure Scenarios

The three major components of an information infrastructure are:

1) electronic networks, including the underlying rules and protocols for their use;
2) databases and knowledge bases connected by networks; and
3) an intelligence (network tools) to allow the users transparent access to databases and knowledge bases.

The infrastructure must build upon current technologies and capabilities, be cost effective, and be capable of evolution (Kahn 1992). The success of an information infrastructure may be measured by the user’s ability to access the right resource, in the right form, at the right time, and at the right cost (Bernstein 1989).

There is no requirement that the infrastructure must offer universal capability to all users. Different users will have different needs that will require different levels of service and sophistication (McGarty 1992). As mentioned earlier, an information infrastructure can provide three classes of spatial resources, or modes of functionality, for spatial data users. First, NSDI can provide spatial datasets and spatial data tools (i.e., software for manipulating datasets) that will allow sophisticated users to solve many diverse and complex problems that may often be unique in nature. The data and the software may be completely decoupled from one another. In some cases, the user may have software, but will need one or more datasets. In other cases, the user may have data but will need software. In still other cases, users will need to find both data and software. To function properly, the infrastructure must allow the user to locate appropriate and compatible datasets and software for their particular applications, yet let the user maintain full control over each individual operation and product. Users at this level will be primarily research scientists and GIS specialists who may provide spatial data products or services. These users will need to know both the value of the data and the appropriateness of the algorithm for their purposes.

Second, an information infrastructure can provide spatial information products (i.e., reports, tables and maps) that users might require to solve many routine problems. These products would be provided by direct or indirect access to spatial database systems coupled with software for manipulation. The users would not fully control operations, i.e., they would use software coupled with the data. However, users would control each product transaction. In most cases, users would be primarily interested in the value and appropriateness of the product and would not be required to have the sophisticated knowledge needed to evaluate the data and software used to derive the product.

Finally, an information infrastructure can provide spatial information services. These services might provide automatic data input to car navigation or air traffic control systems or they might simply be ad hoc services provided by GIS specialists who would deliver spatial information products or who could even provide spatial data tools. Users would have control over the general nature of the services, but would have very little control over individual operations or products. Users would be interested primarily in the value and appropriateness of these services.

Each of these infrastructure functional modes can exist at multiple levels of sophistication. At the highest level, it would employ the full capacity of a network. In its most idealistic form, one could use the infrastructure much as one currently uses a desktop computer or workstation (Smarr and Catlett 1992). Or the infrastructure would provide services to embodied computing mechanisms (Zwart 1993), such as car navigation systems which would receive data through wireless infrastructure transmissions. The workings of the infrastructure would be transparent to the user. A high-level information infrastructure will require a high degree of conformance to standards for formalized metadata to allow system automation (Radermacher 1991). At its lowest level, the infrastructure would employ only the connectivity of resources on the network. While one could still access the system from a desktop computer or workstation under a low-level system, users would require specialized knowledge to navigate the infrastructure. Metadata needs would be less formalized since they would exist primarily for human, not computer, use. Much of the debate on implementing the new National Research and Education Network (NREN) has centered around these issues of capacity (high-end in-
rastructure) versus connectivity (low-end infra-
structure) for increasing usage of the network. The argument
for higher capacity is that it will enable sophisticated ap-
lications that will attract new users. Connectivity sup-
porters counter that simply increasing the number of re-
sources connected to the infrastructure will have the
same effect (Lucky et al. 1992). Most of the growth in the
information infrastructure in the next ten years may
likely be at a low level (Klingenstein 1992).

There is also debate among those who view the NSDI
as a "process" versus those who view the NSDI as a "product." Those arguing that the NSDI should be a
process advocate standards for collecting, processing and
presenting spatial data. They envision common sets of
basic spatial data (also referred to as "core data") be-
ing used across many disciplines and applications. Data
standards, they argue, will ensure consistent measures of
spatial data quality such as accuracy, content and for-
mat. Core data sets will reduce data redundancy and
will ensure consistency between similar data sets, which
would be built from the same core data (NRC 1993;
Coleman and McLaughlin 1990; Jaske 1989). The pri-
mary focus of those who view the NSDI as a process is
the standardization of spatial data.

The counter argument that NSDI should be viewed
as a product concentrates on standardizing how spatial
data may be found and accessed rather than on data
standards. The data standards development process can
be left to market forces rather than to consensus agree-
ments. They argue that mandated standards are difficult
to enforce and that the problems of competing stan-
dards prevent the emergence of truly universal data
They warn that it is dangerous to standardize data since
data collection and modeling techniques are still im-
proving, and that to adopt data standards too soon
would impede new developments (Coleman and
McLaughlin 1990). Standards should, instead, provide
users with tools for describing their data rather than
forcing them to use rigid definitions of data structure
and format (Kuhn 1991). By focusing on the develop-
ment of these tools, rather than on the standardization
of data, those who view the NSDI as a thing believe that
users will gain more choices of spatial data and that
those seeking niche markets in spatial data may be bet-
ter accommodated.

While there are merits to viewing the NSDI as a
process, the author believes that greater long-term bene-
fits may be gained by viewing the NSDI as a thing. It
seems plausible to expect that many different types of
data collection and data processing methods will appear
even in a NSDI with rigid data standards. Powerful
tools for describing these different spatial resources will
help those who wish to find and access certain stan-
dardized spatial resources and to ignore resources in the
NSDI that do not conform to those particular standards.

Much of the new growth in the information infra-
structure is also expected to be commercial (Faulhaber
1992; Kapor and Berman 1992), which will mean that
entrepreneurs will need to develop pricing mechanisms
that adequately recover costs and realize profits in order
to expand services. Current flat-fee pricing structures
for document delivery services (Maxwell 1989) are often
not appropriate for high-end infrastructure applications
that do not entail simple document copying. New price-
ing mechanisms that do not interfere with infrastructure
functionality will be needed (Hemnes 1993). Electronic
information and software are easily copied and distrib-
uted over electronic networks so traditional pricing
strategies may no longer be appropriate.

Some likely scenarios to show how different modes
of functionality might work in a spatial information in-
frastructure are given below. These scenarios view the
NSDI as a thing rather than as a process. They examine
both user and developer issues and briefly discuss how
pricing mechanisms might be used to allow commercial
operations. There will be user costs for the underlying
technology itself—system hookup fees and communica-
tion network usage fees—that may be either added on
to other technology costs as business overhead or
charged separately much as the monthly telephone or ca-
ble television bill. We ignore these costs in our discus-
sions, but they will be present in any possible informa-
tion infrastructure scenario.

The National Digital Spatial Databases System

The National Digital Spatial Databases System
(NDSDBS) is a proposal put forward by the Federal In-
teragency Coordinating Committee on Digital Carto-
graphy in 1989. "The NDSDBS concept is best described as
a system of independently held and maintained federal
digital spatial databases that meet certain minimum
standards" (FICCDC 1989, p. 14). Sets of commonly
used data would be produced to common standards.
Candidate databases would be admitted into the
NDSDBS after technical review of the definition of enti-
ties and attributes and measurement accuracy. Spatial
data interchange within the NDSDBS would be through
the Spatial Data Transfer Standard.

Dataset contribution would be by federal, state and
local government agencies and by the private sector
(NRC 1993; FICCDC 1989). Incentives for non-federal
agencies and private sector data donations would be as-
surance that their data meets rational standards and
that they would be provided a well-recognized venue
for data distribution. It has been suggested that a por-
tion of data collection costs be rebated to donors as an
added incentive (NRC 1993).
The rationale for developing such a system has been stated as follows:

Sets of commonly used data, produced to common standards, will enable users to share data more easily and will allow individual government agencies and companies to concentrate efforts on using, rather than collecting, data. Use of standards also will encourage the development of larger markets and economies of scale, leading to reduced costs for, and increased availability of, data. Dependable supplies of reliable data will increase the Nation’s commerce by giving rise to a new industry that integrates and adds value to geographic data for special markets. (FGDC 1991, p. 3)

A national spatial data infrastructure could be developed as an integrated, distributed system (Jaske 1989) or as a low-level system of separate databases which may or may not be compatible (Onsrud and Rushton 1992).

There are several advantages to developing a national spatial data infrastructure under federal control: First, it would allow for long-term planning of data collection rather than reaction to short-term market conditions (Rhind 1992). Second, there is need to insure the availability of certain data to support research crucial to national stability and security (Galhegi 1992). And third, a federal level infrastructure could provide a framework for developing or improving local government and private sector infrastructure components.

There are several problems with the National Digital Spatial Database System that would need to be overcome. Although it is technically easy to develop usable standards, difficulties arise in trying to convince all the players to agree to standards or to even agree to a process in which standards would be developed. Those supporting proprietary standards have strong incentives to incorporate as many features of their standards as possible into the final version (McCallum 1993). There appears to be no consensus on what data should be included in the infrastructure (Onsrud and Rushton 1992) nor on how that data should be structured (i.e., archives of spatial data sets versus active databases containing the most current spatial information available). These are questions concerning the federal government’s ability to verify donated sets and to update databases and data sets in a timely manner (Mackay and Robinson 1992). And finally, worldwide development of many national spatial data infrastructures, without some agreements among those nations, would still lead to problems with integrating digital spatial data from one national system to another (Salge 1992; Blakemore 1991).

There is also a danger, however remote it may seem to some, that a spatial data infrastructure controlled only by government could lead to the exclusion of data that might be considered politically incorrect or that might call into question official policies and decisions (FGDC 1993). It would seem most likely that a private sector spatial data infrastructure would develop alongside the NDSDBS to provide an outlet for spatial data that would allay these fears and that would better respond to the dictates of the free market.

The user costs for the National Digital Spatial Database System would probably be minimal, since the current U.S. government policy based on the Freedom of Information Act only allows recovery of duplication costs for federally disseminated data. Private enterprises would have little incentive to contribute spatial data to NDSDBS without some means to realize a profit from spatial data transactions. User fees would most likely be based on some flat-fee pricing schedule similar to those already in place in federal government. Collection of fees could be realized through the authorized use of accounts or through a simple credit card system that would match account or credit card numbers against either the customer’s name or against the machine from which the order is being placed. Communications costs for transmitting digital spatial data would be high compared to traditional data because of the large volume of most digital spatial data files.

**Integrated Spatial Data Infrastructure**

The Integrated Spatial Data Infrastructure (ISDI) approach involves attempting to integrate current subsystems into a single, global system. This approach would need to evaluate the problems that may be overcome, incentives for subsystem users to conform to certain communication standards to enable such integration, and possible problems of data inconsistency over such an infrastructure. This approach could include the NDSDBS as well as the many other autonomous systems used by other organizations. It must balance the fears of data owners and collectors who wish to retain control over who accesses data and how it’s used against user fears that data owners will withhold or distort data prejudicial to their commercial viability (Blakemore 1991). This approach would allow users of current systems and networks to slowly and gracefully adapt to using the evolving infrastructure (Kahn 1992). However, it would require a high level of user knowledge or automated metadata to operate efficiently.

Dealing directly with data collectors and owners or their agents rather than through a government controlled spatial data infrastructure would lead to better responsiveness to the needs of data users, including cost-effective updates, availability of new products, and data packaging useful to many classes of users. In many cases, government supplied spatial data would be the de facto standard (Coleman and McLaughlin 1991), but third-party vendors might repackage government data in forms more attractive for commercial users (Lucas
and Rose 1991). Participants would be free to offer data, services, or both.

An Integrated Spatial Data Infrastructure could require translation between many standards and protocols, and while the computing power to do so is available (Mechling 1992), problems of translation due to data semantics will still arise. The advantages of dealing directly with data owners become less tangible in cases where multiple parties contribute to the database. Differing update intervals lead to questions of data currency (Kindrachuk 1992).

User costs for U.S. government-supplied spatial data in the Integrated Spatial Data Infrastructure would probably be the same as under the National Digital Spatial Database scenario. Private enterprise would offer original data or repackage government data at incremental costs. Fees would be collected by automated billing procedures based on user accounts or credit cards matched against the person or machine ordering the data. Private enterprises might require “site licensing” to use the data on more than one machine (Hemnes 1993). Private enterprises might also offer subscription services whereby spatial data is updated at certain intervals (Rhind 1992). Data repackaged into smaller regions or fewer thematic layers would decrease some of the data-transmission costs.

User unfamiliarity with products may inhibit use of some commercial products (Faulhaber 1992). Likewise vendors, particularly those in monopolistic situations, seem reluctant to publish data prices and often prefer to deal with customers on a one-by-one basis (van den Doel 1992).

Distributed Spatial Database Infrastructure

A distributed database management system (DBMS) is a system to manage multiple geographically distributed databases as a single, integrated database. (Mackay and Robinson 1992, p.3)

A Distributed Spatial Database Infrastructure (DSDI) would be based on interconnecting distributed database management systems such that the infrastructure itself resembles a DBMS. The Distributed Spatial Database Infrastructure differs from the Integrated Spatial Data Infrastructure in that the DSNI is based on developing protocols for distributed heterogeneous databases rather than on establishing formats and standards for data exchange. Such protocols are under development for relational database systems but may not be available for object-oriented systems for many years (Lee and McLaughlin 1991). In a Distributed Spatial Database Infrastructure, spatial data would be publicly accessible but spatial analysis applications software would reside within the user’s system.

Distributed spatial database systems can be based on centralized data systems, file-sharing data systems, or client/server (federated) systems (Lee and McLaughlin 1991). A large distributed system would need to be based on a federated approach, allowing both the integration of existing database systems and the extendibility to incorporate new components as they are developed. Databases would be coupled allowing them to interoperate by exchanging information and providing services for one another (Mackay and Robinson 1992; Heiler 1989). Tightly coupled systems require good central control mechanisms, but provide better consistency and performance. Changes to one component could dramatically affect other parts of the system (Heiler 1989). Such systems could be viewed as a virtual database or as a seamless GIS with users unaware that data is being retrieved from different sources (Baker and Broadhead 1992; McAbee 1992; Lee and McLaughlin 1991). Loosely coupled systems allow more autonomy among database components. They are better for dealing with data complexity, are easier to maintain, and are often more responsive to user needs, since the user, not a central control mechanism, would control interactions between components. However, loosely coupled systems require more user knowledge, will result in more data inconsistencies, and may encounter performance problems due to excessive data translations (Heiler 1989).

The greatest advantage of DSNI would be that users need only retrieve those portions of datasets actually needed rather than access whole datasets and reprocess them to meet actual needs. Data filtering would be especially necessary and advantageous in a free market economy where users would only pay for what they actually need and use.

Another cited advantage of distributed spatial database systems is the elimination of data exchange standards (Mackay and Robinson 1992). However, there is still a need to develop common, global data models that can be mapped to the local schema of all members of the infrastructure (Baker and Broadhead 1992). Development of global data models is hampered primarily by the problem of representing model constraints of local schema. Translating a local schema to a global model and then to another local schema gives rise to all the problems inherent in data exchange, particularly the loss of representation and modeling power (Kim and Seo 1991; Radermacher 1991; Worboys and Deen 1991).

There are some disadvantages to DSNI. Problems occur when one system misinterprets data from another system or when incorrect services are provided. The causes of such problems are generally semantic in nature. Semantic data models are often inadequate at describing the behavior of objects in local schema. Differ-
ent names may identify the same concept or the same name may identify different concepts among local schema. Concepts may be related through different dependencies among member schema (Mackay and Robinson 1992; Heiler 1989).

Data would still need to be replicated in a Distributed Spatial Database Infrastructure to provide data security against accidental loss of data and to allow dependable access to data in the case of one data site being temporarily out of service. The need for data duplication will require strict multi-site updating and concurrency control procedures (Lee and McLaughlin 1991).

From a purely organizational standpoint, it is difficult to imagine that the necessary agreements could be made to operate a large portion of the information infrastructure as a single distributed database infrastructure (Faulhaber 1992). Differences in organizational goals and operating procedures would seem to strongly prohibit lasting agreements. There are also difficulties defining the boundaries that constitute "spatial data" that could allow individual systems be included in the Distributed Spatial Database Infrastructure.

User fees in a Distributed Spatial Data Infrastructure could still be collected through automated fee collection of user accounts or credit cards, but billing procedures become a bit more complicated. If users are allowed to filter data, then flat-fee pricing would probably give way to either data-unit pricing or connection time fees or both. Users could tie up databases attempting to minimize data unit costs and be subject to additional connection time costs. There could be some initial difficulty establishing reasonable data unit prices, since these prices should reflect data quality; but once the initial difficulties are overcome users would have a much more uniform method of comparing data prices. However, some land information systems have already established data-unit pricing based on the land ownership parcel (Kozub 1992; McKay 1992; Merrick and McKay 1992; Scrivens 1992).

**Spatial Metacomputing Infrastructure**

The Spatial Metacomputing Infrastructure (SMI) approach is based on metacomputer research and prototype development at the National Center for Supercomputing Applications (NCSA). A Spatial Metacomputing Infrastructure would consist of a "network of heterogeneous, computational resources linked by software in such a way that they can be used as easily as a personal computer," (Smarr and Catlett 1992, p. 44). Such an approach would allow users to attempt spatial analysis and other applications that would be nearly impossible without a metacomputer. This approach differs from a Distributed Database Information Infrastructure in that distributed spatial data and applications to process such data would both be included in the infrastructure. Software capable of handling such computations are now emerging, although telecommunication speeds still lag for most applications (Smarr and Catlett 1992).

A Spatial Metacomputing Infrastructure would seem to have the advantage of allowing easy integration of spatial data into the general information infrastructure, allowing spatial applications to be reliably transferred to desktop publishing tools (McAbee 1992; McLaughlin 1991). Development of a SMI would bring all the problems of developing a Distributed Spatial Database Infrastructure with the added condition that not only must spatial databases be compatible, but also spatial analysis packages be compatible to both one another and with other non-spatial infrastructure applications. Users will need to be able to specify well-structured classes of computations that could be implemented with relative ease (Kahn 1992).

There would likely be some resistance from software developers, since users would no longer need to purchase software. Revenue collection based on monitoring the use of software, rather than on software distribution, could be one solution to overcoming such resistance (Hennes 1993; Cox 1992). However, public domain geographic analysis tools are available and some organizations are developing public domain software specifically for network use. For example, the NCSA Software Tools Group (STG) is developing public domain data analysis and visualization software useful for many non-geographic spatial data purposes (i.e., molecular modeling), some of which appear to be adaptable for geographic purposes.

User fees in a Spatial Metacomputing Infrastructure would probably be based on data-unit pricing plus metered software use. In some cases, this could result in two bills, similar to separate local and long-distance telephone bills. Automated bill collection based on user accounts or credit cards could still suffice for ordering data while monitoring hardware connected to computers running software applications could keep track of software usage and report the amount of usage at weekly or monthly intervals to a billing company.

**Embodied Spatial Infrastructure**

The Embodied Spatial Infrastructure (ESI) surmises that spatial data and applications would become "embodied," or transparently integrated, into other enabling technologies (Zwart 1993). Services could include support for automobile navigation and air-traffic control systems (Wegener 1993) or the infrastructure could allow users to specify a product for a specific purpose without every realizing the spatial data origins of that product (NRC 1993). This concept is actually an extension of the Spatial Metacomputing Infrastructure, except
that applications would be transparent. Some ESI applications would require a limited set of data types with a well-defined set of operations available in the public domain (Zwart 1992). Other applications would require strong market agreements between entrepreneurs providing competing or complementing services. At the support level, spatial data specialists providing embodied services would still need to access appropriate spatial data. The timeliness of spatial data updating may become a more critical concern, particularly in applications where human lives are put at risk. For example, an ambulance service relying on an automated navigation-updating service would need up-to-the-minute road and traffic conditions in order to optimize routes between accident sites and hospitals.

An Embodied Spatial Infrastructure would be able to handle narrowly defined, routine operations that meet well-defined, seldom-changing tasks but would not be appropriate for many unique or evolving applications (Wegener 1993). It will require a better understanding of spatial concepts than are currently known in order to provide many of the transparency mechanisms necessary for implementation. Since ESI applications would be narrowly defined, the data semantics problems of DSDI and SMI would not be as acute nor as problematic.

User costs in an Embodied Spatial Infrastructure would probably be simple. Fees could be flat-rate structured, similar to current cable television subscriptions, or fees could be based on metered use. In certain cases, such as local chamber of commerce tourist-navigation information, services might be free to users (paid by local businesses).

**Summary of Spatial Data Infrastructure Scenarios**

There are five possible scenarios for a spatial data infrastructure that have been suggested in the current literature and which have been expanded and examined in this paper. While these scenarios have been treated as distinct and separate entities, in reality they overlap and share common components. These scenarios have differing mixtures of archived data sets, active databases, and analytical software as their primary components. A summary of these components is shown in Table 1.

It seems likely that several of the infrastructure scenarios described may come about. Each seems to fit a special niche or to provide a simple way of extending existing spatial resources in the infrastructure. Spatial resources in the National Information Infrastructure may exist as described in one or more scenarios and gradually evolve to one or more higher level scenarios. New scenarios may develop as we learn more about the possibilities of infrastructure architecture design. A summary of our scenarios can be found in Table 2.

In planning for the future, we must be careful not to foreclose options that may lead to better or more extensive uses of the infrastructure (White House 1993; Meckling 1992). Nor should we force the user to abruptly relearn the system each time it changes or new resources types are introduced. We should plan for the evolution of the infrastructure to higher levels, but realize that the evolutionary process may be uneven and that the results may not always be those that we have envisioned. Accordingly, navigational paradigms should recognize these infrastructure design strictures.

**Infrastructure Navigation Paradigms**

The infrastructure scenarios described above all require navigational aides. The usefulness of spatial data, software, products and services are all difficult to assess without external information (metadata) that can describe at least a minimal set of parameters that can be used not only to evaluate the resource, but also as a set of keys to narrow the scope of the search. The most obvious solution is to look to cataloging practices to consolidate and extend this needed external information. Cataloging paradigms answer three important user questions:

1. Does it exist?
2. What exactly is it? and
3. How can I get it?

<table>
<thead>
<tr>
<th>TABLE 1. Components of Proposed NSDI Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National Digital Spatial Databases System (NDSDBS)</strong></td>
</tr>
<tr>
<td>Archival data sets</td>
</tr>
<tr>
<td>Active databases</td>
</tr>
<tr>
<td>Analysis software</td>
</tr>
<tr>
<td>Spatial services</td>
</tr>
</tbody>
</table>

*Frank URISA Journal 45*
TABLE 2. Summary of Spatial Data Infrastructure Scenarios

<table>
<thead>
<tr>
<th>Characteristics:</th>
<th>National Digital Spatial Data Base System (NDSDBS)</th>
<th>Integrated Spatial Data Infrastructure (ISDI)</th>
<th>Distributed Spatial Data Infrastructure (DSDI)</th>
<th>Spatial Metz-computing Infrastructure (SMI)</th>
<th>Embodied Spatial Infrastructure (ESI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-end applications</td>
<td>Low-end applications</td>
<td>High-end applications</td>
<td>High-end applications</td>
<td>High-end applications</td>
<td>High-end applications</td>
</tr>
<tr>
<td>High user knowledge</td>
<td>High user knowledge</td>
<td>Low user knowledge</td>
<td>Low user knowledge</td>
<td>Low user knowledge</td>
<td>Low user knowledge</td>
</tr>
<tr>
<td>Federal government distributes federal and contributed data</td>
<td>Integrates current and new spatial data sources</td>
<td>Integrates current and new spatial databases</td>
<td>Integrates databases and analysis software</td>
<td>Embodies spatial data and software into other technologies</td>
<td></td>
</tr>
<tr>
<td>Advantages:</td>
<td>Government vetted datasets</td>
<td>Deal directly with data supplier</td>
<td>Common operational protocols</td>
<td>Provides many routine services</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allows long-term planning</td>
<td>Commercial re-packaging of govt data</td>
<td>Retrieve only portions of dataset needed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disadvantages:</td>
<td>No consensus on which data to include</td>
<td>Differing data and protocol standards</td>
<td>Data semantic problems</td>
<td>Need to make applications compatible</td>
<td>Unable to handle custom needs</td>
</tr>
<tr>
<td></td>
<td>Possible exclusion of data for political reasons</td>
<td>Difficulty obtaining organizational agreements</td>
<td>Need to develop new software use cost structures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>User Costs:</td>
<td>Flat-fee copying cost pricing</td>
<td>Flat-fee document pricing</td>
<td>Spatial data unit pricing</td>
<td>Flat-fee or spatial unit pricing plus metered software usage</td>
<td></td>
</tr>
</tbody>
</table>

The first question is answered by indexing, the second and third questions are answered by the catalog, or metadata, entry.

Spatial data is often best indexed by geographic location. Users searching for information about a certain geographic area can retrieve all the information for that particular location if it is spatially indexed without having to know anything other than the location. If no geographic location "key" is provided (i.e., geographic coordinates, place name, street address, etc.), users must resort to more clumsy retrieval methods and hope that all pertinent information may be found. Spatial indexing is a feature of nearly all spatial data collections and differentiates them from the vast quantities of resources that are not spatially referenced. Many of these spatial resource systems have developed proprietary graphical interfaces that allow users to specify spatial location on a map representation shown on the screen. These interfaces require client software that can communicate with the server and are not interchangeable. Network retrieval protocols based on spatial indexing will be necessary in order to build interoperable client-server tools that will allow users to effectively locate spatial data in the information infrastructure.

The catalog, or metadata, entry serves as a surrogate for the document, dataset, database, software or service. The catalog entry can exist either with the document, dataset, database, software or service itself, or it can reside separately within metadatabases. The arguments against separating metadata used for cataloging from the source are that much of this metadata is common to the metadata needed to operate the source material or service and that concurrency problems might arise when this metadata at the source is updated. The argument for separating the two are that the source might not be accessible and that separate catalog metadata records can be easily assembled in specialized information catalogs. Those arguing for separation of metadata also note that much metadata is common to many datasets and including it in each dataset would be an unnecessary duplication effort.

The user of a network of resources will need to view that network as both a collection of entities and as a single entity in its own right. At times, the user may wish to retrieve or search for spatial resources from a single, known network site. At other times, a user may wish to search or browse for spatial resources at selected multiple network sites. And finally, the user might wish to browse or scan for spatial resources all across the infrastructure. Navigational paradigms should allow the user to decide where and how to find a given resource.

The navigational strategies outlined below are based on current resource location system (RLS) mechanisms,
using the client-server model inherent in many of those mechanisms. They postulate how such mechanisms could be extended to locate spatial data in the information infrastructure and evaluate the value of such approaches to spatial resource users.

**Super Catalogs**

One possible strategy would be to create a distributed "super" catalog that would contain descriptions of spatial resources with pointers to the actual servers containing a given resource. This concept would be modeled after the Wide-Area Information Server (WAIS) "Directory of Servers" (Kahle and Medlar 1991). The user would first direct his or her query to the super catalog, which would then point to a number of possible resources that might meet the user’s needs. The user would select promising-looking resources and redirect the query to that resource.

For users wishing to find servers containing information about certain geographic areas, this would require that the descriptions in the super catalogs include relevant spatial information—such as coverage, scale, data formats, etc.—necessary to narrow down their searches. It would also need to include information about the particular catalog or information services capabilities. What is the organization of the system? Can the user apply temporal constraints? What types of data or information are held? There may be difficulties in abstracting certain resources because of the multiplicity of holdings within those resources. For example, an on-line spatial data catalog containing all the varied spatial data holdings of an organization would need a complex super catalog description to cover all those holdings.

Once the user is dropped from the super catalog to the appropriate resource (which might itself be a metadata database or catalog), he or she will be required to understand the logic of the system and the organization of the interface in order to effectively use the system. As systems continue to proliferate, this may require a continual learning of new system organizational strategies and interfaces. Alternatively, there may be a narrowing of system organizational designs and interfaces as the infrastructure stabilizes. However, it may take some time before this actually happens. The super catalog, without the accompanying standard services offered by a system like WAIS, is merely an automation and refinement of the serial access FTP and TELNET services. Users would still need to learn and remember how to retrieve, search for, browse, or scan each system that might contain relevant information for their needs. The high level of data abstraction needed to describe many different types of spatial resources could make data searching more difficult, since it may be difficult to ascertain from individual super catalog descriptions whether a re-

source actually contains or lists the data or software that a user wishes to access.

The super catalog concept might be extended to include spatial resources held within particular on-line resources, which are cross-indexed with one another. Thus, a series of super catalogs, each with a different specialty, might index different types of resources. These specialty super catalogs could then be collected and placed in a master super catalog.

Low-end infrastructure super catalog applications, such as the National Digital Spatial Database System and the Integrated Spatial Database Infrastructure, make no demands that the spatial resource listed be available on-line. In fact, many spatial resource catalog systems now operating list many spatial resources that are not available on-line, but must be obtained off-line through direct agreements with the resource producer. Low-end applications still suffer from myriad user interfaces that may be encountered when dropped from the super catalog level to the spatial resource level. Low-end applications, primarily for human and not computer consumption, would encourage human updating at many sites. However, this updating, because of the human element, would likely be inconsistent due to differing cataloging priorities and preferences.

Higher-end infrastructure applications, such as the Distributed Spatial Database Infrastructure, the Spatial Metacomputing Infrastructure, and the Embodied Spatial Infrastructure, will require that all listed resources be available on-line. They will also require that extensive spatial metadata, particularly data dictionaries, be attached at the super catalog level to allow applications to determine resource compatibility. They would need to contain additional information to allow the user to find known compatible resources. If these metadata extensions are not done to common standards, many resources will have only limited interoperability with other systems. The inconsistencies of human input would also likely hamper interoperability between many systems. There appear to be no incentives, other than a desire for interoperability, to persuade users to conform to common standards in a super catalog navigation system.

There is one implementation of a super catalog in use and another being proposed. The WAIS "Directory of Servers" is a server containing descriptions of WAIS servers (Kahle and Medlar 1991). The user first directs a query to the Directory of Servers that will return a list of possible servers. The user then redirects his or her query to one or more of the listed servers. The WAIS implementation of a super catalog is simple, requiring only that one set up a WAIS server and register it with the Directory of Servers. And it is homogeneous, allowing the user to query the Directory of Servers and individual WAIS servers with the same interface. However,
WAIS is not very flexible, requiring that users search under the WAIS protocol and not use other methods. In practice, one must often query the Directory of Servers using broad terms, then refine those terms with the individual servers. Other searching methods could be added to the WAIS protocol—USGS has added spatial searching capabilities to WAIS (Nebert 1993)—but users would still be restricted to searching under the methods supported by WAIS. Nor is WAIS extensible. The Directory of Servers only lists WAIS servers and does not list other types of servers.

The Government Information Locator Service (GILS) is designed to catalog information resources within the U.S. federal government (Christian 1993; McClure et al. 1991). It will list and describe inventories, catalogs, databases, clearinghouses and offices where federal government information might be found and accessed. A "GILS Core" will aggregate resource information to approximately 1,000 entries, with at least one entry for each major source of information. Topical directories will differentiate information types (Christian 1993). GILS will be extensible, allowing all new federal government resources to be listed. It is unclear whether that extensibility would include local government or private resources. GILS is simple, allowing users to begin searches at a high level, although the problem of beginning with broad terms and later refining these terms will probably still exist. There seems to be a tradeoff between flexibility and homogeneity in GILS. If users may search using the different interfaces associated with individual resources, homogeneity will be sacrificed. But users would still be restricted to searching in a given resource by the individual system interfaces. If users are required to use a common interface, such as WAIS, flexibility will be sacrificed. Overall, super catalogs have low flexibility, medium extensibility, high simplicity and medium homogeneity.

**Virtual Catalog Cards**

The virtual catalog card (VCC) concept rests on adopting or developing a standard spatial metadata exchange format. This format would be used by spatial cataloging or information systems to request and respond to search queries. Exchange formats are generally field oriented with an acceptable range of values for each field. The client interface would receive a request from a user, which would then be composed into a standard exchange format record and sent to one or more servers for a response. The server would then map the contents of the exchange format record to its internal structure to respond to the query.

One advantage of a standard exchange format would be a clear delineation of entry points for searching for data. A spatial metadata exchange format might have fields for identifying data source, data scale or resolution, data publisher, data extent and data cost. Clients and servers exchanging information would need to map the appropriate exchange field to their own internal formats. In the case of free text descriptive material, this may entail more sophisticated methods than mere field-matching techniques.

A drawback of VCC is that retrieval, searching, browsing and scanning for information might still be dependent on users understanding the organization of the resource they are accessing, since these resources could only map compatible access points to the exchange format. Also, client applications may be left to interpret empty format fields returned from servers as best as they can. An empty field may mean that the information value is nonexistent in the server system or that an acceptable value cannot be mapped from the server system. The latter might occur if the field is not a point of entry for accessing records in the server system.

Adoption of a standard exchange format will require strong agreements on conformance if interoperability is to be achieved (Lynch 1993). A virtual card catalog record for low-end infrastructure navigation could use existing exchange formats. Cataloging exchange format used by the library community. The DIF format allows different users wide descriptive materials that are suitable for human consumption but could be difficult to map consistently to exchange formats. MachineReadable Catalog has been used successfully by libraries, but has shown some deficiencies when used for spatial data (Larsgaard 1992; Lai and Gillies 1991; Holmes 1990). However, at least one apparently successful spatial data catalog, the Louisiana Coast GIS Network (LCGISN), is MARC-compatible (Hiland, Wayne, and Streiffer 1992; McBride et al. 1991). A low-end VCC would be easy to implement and could be easily learned by those accustomed to using existing systems.

Since results returned from the server would also be mapped to the exchange format for display by the client, users would be able to develop a single, customized interface to deal with a single resource location searches. Like the super catalog, there is no requirement that the spatial resources in the low-end infrastructure listed in the VCC record be available online.

A virtual catalog card record for high-end infrastructure applications, including the Distributed Spatial Database Infrastructure, would require that the resource listed be available on-line and would involve complex record structures that would need to map data dictionaries and other high-level operating information. Extending a low-end approach to high-end applications will also require a good degree of cooperation and understanding. However, these traits have not always been present in current exchange standard applications (Lynch 1993). Automated means to abstract the needed
VCC record information directly from spatial resources would be highly desirable.

User flexibility under VCC records would be medium, since users would not necessarily be allowed to search on the VCC record elements within different catalog systems, but only on those elements indexed by the server system. Clients would, however, be able to refine the records returned from those systems based on the individual elements. Extensibility would be possible, but would likely be a slow and drawn-out process as new resource types are incorporated into the standard. Extensibility will also be rated as medium. VCC records will be quite complex to accommodate the myriad type of information to be searched for and accessed. Homogeneity would be high, since client interfaces using the VCC record format could be constructed according to individual user tastes.

Hyperlinked Spatial Resources

Incorporating metadata at the source will require tools to easily extract it for analysis and use. For documents and datasets, this could be done by structuring the file sources using Standard General Markup Languages (SGML), which uses descriptive tags to separate the logical elements of a data or document file. These tags allow mechanisms for searching through the files, and for displaying selected portions of the file (Adler 1992; Kiser 1989). The SGML is defined in ISO standard 8879 (Hajagos 1993). It is not itself an application language, but a modeling language allowing different applications to be identified and processed independently. Use of SGML requires significant preliminary work to define application requirements and then design a suitable application (Adler 1992).

SGML structures contain three main elements:

1) SGML declaration;
2) document type definition (DTD); and
3) the document instance.

The declaration contains information about the encoding methodology, base character set used, symbols used for tag descriptors, and the maximum length of tag names. The DTD defines the structure and context of the document. The DTD also describes relationships between tags. The document instance is an actual marked up data or document file described by the declaration and DTD. The declaration and DTD can reside in the header of a document instance or may be referred to in the document instance header. Several document instances may refer to the same declaration and DTD. In a well-defined SGML application, the SGML syntax is transparent to the user (Hajagos 1993). The most visible network application of SGML is with the World-Wide Web (WWW). The U.S. Geological Survey's Global Land Information System (GLIS) on-line catalog makes use of a simple SGML schema to add information to dataset descriptions. The use of SGML has been put forward as an alternative to current library cataloging techniques (Hirshon 1993).

The SGML standard currently addresses only textual information (Hajagos 1993). In a spatial data infrastructure, data files would be linked at creation to a file containing an index of datasets and to other pertinent data and document files. These other files might contain parent data (data used to derive the current dataset), data collection and processing methodology, data quality reports, and products—such as reports, tables, and maps—generated from the dataset. Any dataset updating would require either concurrent updating of affected hyperlinked files or the creation of new hyperlinked files. There is a trade-off in user and supplier convenience in using hyperlinks. Suppliers face the danger of being seduced by the ease of creating a new hyperlink file for every update rather than locating appropriate linked files and updating them also (which may require other files to be updated), thus requiring searchers to visit many more files in determining the fitness of a dataset.

A hyperlinked spatial data infrastructure coupled with searching protocols could allow users to perform Boolean searches based on tags. For example, a tag specifying data scale and a tag describing thematic or topical content could be combined to find all 1:10,000 datasets with vegetation coverage. Additional tags would provide pointers to link the dataset or document to related datasets and documents. The user would be able to browse through related datasets by exploring the links in the retrieved data or document files. To be effective for spatial data searching, the searching protocols would require a mechanism for performing spatially indexed searches. There seems to be no actual requirement that all data and document files in the infrastructure adhere to the same SGML definition. However, files that are hyperlinked to one another may need to have identical or highly compatible definitions.

If only spatial metadata are hyperlinked spatial resources need not be available on-line. Hyperlinking spatial metadata directly to spatial resources could eliminate the need for duplicate metadata residing in both catalogs and within the resources themselves. Hyperlink extensions could easily be added to enable higher-end applications without the constant need to hammer out new agreements at each intervening level, as might be necessary with super catalogs or virtual catalog cards. However, since the SGML definitions only apply to text-based materials, there may be great difficulty hyperlinking non-text spatial resources such as software. These resources could be linked from a dataset file to
the resource, but not among one another as would be needed for high-end infrastructure applications.

Construction of hypertext can double the time needed to input data—few automated tools for doing so are available (Stephenson 1988). Critics of hyperlink claim that navigating these links can leave users “lost in hyperspace” (Shepherd 1991, p. 355), unable to return to their point of origin. Another deficiency of hyperlinked spatial data files is that the files may not be compressed to preserve archival space (or if the files are compressed, they must be uncompressed to search for the hyperlinks). And whenever files are removed from network accessible locations, the pointers to those files must be edited from all the files that call to it. Embedding hyperlinks within a resource also presumes that the information necessary to operate the resource is readily available within the resource itself, which is rarely true for many spatial resources. Users will not have the time nor the patience to continually access a manual, even if by hyperlinked pointers, to determine how to proceed within a given resource.

Finally, the success of hyperlinked files will be determined by the search strategies used to gain that important first link. If users are required to browse numerous sources or wait while search engines examine thousands or more documents at multiple sites, any advantages of hyperlinking data will be lost to the huge up-front time needed to find the appropriate initial link.

Users would have tremendous browsing capabilities using hyperlinked information. However, they would be required to remember or mark links in order to retrieve known data. Hyperlinking has poor searching capabilities which would need to be overcome by offering an alternative searching mechanism. Those scanning for data would need to do much work as they follow different paths looking for interesting items.

Hyperlink schemes, which are conceptually little different from hierarchical indexing schemes, seem well-suited for both low-level and high-level infrastructure applications, but would seem to limit interoperability to resources which are linked to one another. Users would be required to use the hyperlinks, and would thus have a low level of flexibility. Extensions could be easily added, but are currently confined to textual materials, and should be considered medium. Hyperlinks appear to be difficult and complex to build and to maintain. Users would see a moderate level of homogeneity. They would be able to constantly use hyperlinks, but perhaps would see these links used differently from system to system.

**Metadata Collectors**

A series of specialized, distributed network servers could be used to collect spatial metadata, thus limiting users initial searches to a limited number of specialized resource location systems (RLS) (Duestch 1992). Such RLS, patterned after the Internet archive servers, would contain spatial metadata and pointers to the actual data. One or more RLS servers would be used to either proactively or reactively gather metadata, including the unique resource number (URN) and universal resource locations (URL) for the actual resource (Lynch 1993). In a proactive state, the servers would access network resources, inquire if the resource fits the particular specialization of the server, and retrieve the metadata. Retrieved metadata would be compared against existing server metadata and the server updated only when changes are found. In a reactive state, the resource would broadcast any changes to the appropriate server, which would then update its information accordingly (Duestch 1992).

The organization of these RLS could be by discipline, by geographic location, or by data type. Disciplinary resource location systems would gather metadata for spatial data considered useful to certain disciplines. Disciplinary systems would overlap with other disciplinary systems when spatial data is considered useful to more than one discipline. Geographic location resource location systems would gather metadata for predefined geographic areas such as Maine, the United States, or North America. Geographic location systems likewise could contain much redundant metadata, the amount depending on the hierarchy of the geographical location classification. However, geographic location systems could be hierarchically linked to avoid redundancies. Data type resource location systems would gather metadata for certain types of spatial data (i.e., vector datasets, raster datasets and imagery). Data type systems would contain less redundant metadata. They would also be easier to organize and to search since they only need to include parameters relevant to the data type contained in the system.

Problems would arise if there was no agreement on the organization of RLS. Users, or implementors of client applications, would be required to learn different search and browse strategies for each type of RLS—one strategy for disciplinary RLS, another strategy for data type RLS, and still another strategy for geographic location RLS.

The metadata could be collected from on-line data files, using a process such as the virtual catalog card record or SGML to mark relevant items for collection, or could be supplied in separate files which would contain series of metadata listings for a given resource. In the case of software or services, metadata would have to be placed in a file available at the source.

Users may still need to learn many organizational strategies if they wish to search more than one collector system. Retrieving, searching, browsing or scanning
would be effectively narrowed down in cases where the collector systems reflected the user’s constraints. In other cases, users would need to develop strategies, then apply them to the multiple collector systems.

Again, low-level metadata collectors would not require that the spatial resource itself be available on-line, only that a description of the resource available. High-end applications will require that the resource be available on-line. Since much of the resource information is collected automatically, extending low-level applications to higher levels should be relatively simple compared to super catalog, virtual catalog card, and hyperlinked approaches.

Users would have medium flexibility, limited to searching on the terms collected by a given metadata collector system as well as with the interface to that individual system, but perhaps able to search on different sets of metadata collectors using different metadata collection strategies. Extensibility would be medium. Finding the proper collector to collect metadata from a given resource may be problematic, but new metadata collectors could be added as needed. Metadata collectors are themselves simple, but finding plausible strategies to organize metadata collection strategies may be difficult. Homogeneity would be medium, assuming that a limited number of metadata collector types would be built and that these systems would share portions of their interfaces.

**Service Protocols**

Service protocols would allow clients to query servers as to the nature and extent of their services. For example, a spatial resource server containing spatial metadata might respond to an initial client query that spatial resources are available. It might go on to list how the metadatabase may be searched, i.e., by spatial constraints based on coordinates, by keyword, by data collection method, by data scale or resolution, etc. Each of these search mechanisms would be associated with an appropriate service protocol that would describe how that search would be presented to the user. Similarly, a resource containing software for analyzing spatial data might respond to an initial query that data processing services are available. It might then list the range of operations possible. Each operation would have an appropriate service protocol detailing how the data must be presented to the software and how the results would be presented to the user.

In the client-server environment, the same results from a single server could look different from client application to client application, but the actual result content from the server would be identical. Similarly, different servers would provide different implementations or levels of a given service, but would present the results in an identical format.

Service protocols would not work by themselves. They would need to be combined with something like a virtual catalog card record to broadcast what services are available. They might also need to have some type of overall controlling mechanism, such as a super catalog or metadata collector, to allow high-level searches for services.

Service protocols would work well at both low and high levels of the infrastructure, since they apply concepts and not merely mechanisms and thus may be implemented in a variety of ways, as long as they exhibit similar behavior. The extension from low-level to higher-level applications should be more straightforward, since one would only need to extend the characteristics of behavior, and not unduly worry about the methods to be used to cause this behavior. Chances for interoperability between systems seems greatest with a suite of service protocols, since different systems would be free to upgrade or add new services without disrupting existing interoperability levels. However, service protocols will themselves take time to develop and implement, particularly in the voluntary atmosphere of the current information infrastructure standards development environment where a proposed standard must remain open to comments for six months before it can advance to a draft standard. Draft standards require another four months before they can be promoted to standards (IAB 1992).

Service protocols are highly flexible, since they specify how a given resource must accept and present information, not how they must implement the protocol. Users would be able to search and access resources under any client navigation system that incorporated the same protocol or set of protocols as the server system. Clients and servers could incorporate different protocols that seem to accomplish the same task, thus increasing their flexibility. Service protocols also seem to be highly extensible. New protocols could be developed as needed for new resources or for new methods of searching and accessing existing resources. If developed one by one, service protocols would be simple, each protocol defining a single concept needed to allow a specific task. Service protocols would promote homogeneity, since a single client interface could incorporate all the protocols necessary for a specific set of tasks.

**Summary of Infrastructure Navigation Paradigms**

No single navigational strategy seems to offer a simple and effective way to allow users with widely differing needs to find and access spatial resources with universal application. The choice for future development seems to be to create multiple subsystems for spatial resources. For instance, different super catalogs or metadata collec-
TABLE 3. Comparison of Cataloging Paradigms

<table>
<thead>
<tr>
<th></th>
<th>Super Catalogs</th>
<th>Virtual Catalog Cards</th>
<th>Hyperlinked Spatial Metadata</th>
<th>Metadata Collectors</th>
<th>Service Protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexibility</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Extensibility</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Simplicity</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
</tr>
</tbody>
</table>

tors would be developed for different disciplines or applications with little or no system interoperability. Universal transfer or service mechanisms describing how different spatial navigation operations (i.e., retrieval, searching, browsing) would behave would be developed to provide the needed interoperability. These mechanisms must be free of technological restraints, although some application differences may be tolerable. A summary of the rankings for the different cataloging paradigms is given in Table 3.

It seems likely that we may need to combine several of the described approaches to provide efficient navigation at both low and high levels of the information infrastructure since parts of the infrastructure will likely exist concurrently at many different levels. For example, super catalogs could list a wide variety of resources, including hyperlinked spatial resources and metadata collectors. Virtual catalog card records could be used to forward requests from the super catalogs to the resources it lists using service protocols collect the elements in the VCC record. Similarly, metadata collectors could serve super catalogs, collecting and aggregating information about resources.

TABLE 4. Summary of Spatial Data Infrastructure Navigation Mechanisms

<table>
<thead>
<tr>
<th></th>
<th>Super Catalogs</th>
<th>Virtual Catalog Cards</th>
<th>Hyperlinked Spatial Metadata</th>
<th>Spatial Metadata Collectors</th>
<th>Service Protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics:</td>
<td>Lists spatial resources</td>
<td>Spatial metadata exchange format</td>
<td>Uses Standard General Markup Languages (SGML)</td>
<td>Proactively or reactively collects spatial metadata</td>
<td>Allow compatible services between different systems</td>
</tr>
<tr>
<td>Hierarchical system dropping uses into actual resource</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advantages:</td>
<td>Few initial sources to contact</td>
<td>Defines entry points for spatial searches</td>
<td>Allows active links between spatial resources</td>
<td>Different collectors can specialize in different types of spatial resources</td>
<td>Allows many different approaches to be compatible</td>
</tr>
<tr>
<td>Provides exchange medium for information</td>
<td>Strong browsing capabilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disadvantages:</td>
<td>High level of abstraction can cause uncertainty if resource meets user needs</td>
<td>Different spatial data can be searched differently</td>
<td>SGML currently handles only text</td>
<td>May require different search strategies for different collectors</td>
<td>May require combination with other approaches</td>
</tr>
<tr>
<td>User may be confronted with many different resource level menus</td>
<td>Requires strong agreements among cataloging systems</td>
<td>Difficulty in maintaining links</td>
<td>Different SGMLs may be incompatible</td>
<td>Protocol development process takes time</td>
<td></td>
</tr>
</tbody>
</table>
Metadata collectors could also collect metadata from hyperlinked spatial resources, using either service protocols or VCC records which incorporate service protocols. Service protocols seem to be a common thread linking together with the other various navigation strategies. Service protocols also are a good starting point for allowing interoperability within and between the various navigation strategies. The author believes that service protocols should be developed and used to describe the theoretical workings of infrastructure tools necessary for spatial resources. These protocols would eventually become extensible to cover the behavior of many spatial navigation tools. Those protocols common to other infrastructure resources should also be adopted and used by the spatial resource community. Table 4 contains the summary of the characteristics of the different navigation mechanisms presented in this paper.

Conclusions

The ability to quickly and easily locate and access appropriate spatial resources will be the foundation for success of those resources within the National Information Infrastructure. Many infrastructure tools are available or under development that show great potential for needs of spatial resource users in the NII. However, no one tool seems to provide universal methods of finding data suitable to the variety of ways that data can be found. Developments in system interoperability may eventually allow users to transparently switch approaches and applications as their navigation needs change in the pursuit of spatial resources.

Some tools, specific to the spatial data community, will need to be developed. While universal resource numbers and universal resource locators show a promise to solve many problems, spatial resource users will also need universal methods to locate resources based on geographic and/or temporal indexing methods and universal methods for assessing the usefulness of spatial resources. While the universal location methods can be solved in the short term with the development of geographic and temporal retrieval protocols, methods for extensively standardizing assessment evaluations (i.e., how the data was collected and processed) may be impossible due to the varying nature of different spatial resources.

References


References


Features

In This Issue...

To establish the appropriate mindset for URISA '94 in Milwaukee, we offer an array of information technologies (IT) and their varied applications that make "sense" in the Great Lakes region.

Anyone who lives along a coastline, be it ocean or lake, is aware of the unpredictable nature of water and its erosive forces. Carol Johnston and James Salés use GIS to trace the history and plot the future of erosion hazards along the Lake Superior shoreline.

Michael Beaulac and co-authors trace the evolution of MIRIS, Michigan's Resource Information System. In partnership with both the public and private sector, MIRIS represents a user-driven, marketing approach to management and data-layer development that has fostered trust in its products and services.

GIS warfare? Not exactly, but it is a powerful weapon in the often highly politicized process of congressional redistricting. Ken Strasma and Brenda Haskins take us through the "ways and means" of the redistricting process in Wisconsin following the 1990 census.

Will Craig outlines the rising use of GIS technology in Minnesota's state and local government and the associated problems in data management and coordination. He looks at several illuminating examples and offers recommendations for heading off some of the problems.

Back to water, and Lake Michigan in particular, Joel Mouradian and co-authors describe how they integrate remote-sensed and on-site data to study water quality and provide a better understanding of lake ecosystems.

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 fifth year, this conceptual framework appears to be working. We will, accordingly, carry on.

In the Feature 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!—contributing material to the Features section is pleasurable. So, please do so.

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

Warren Ferguson
Lynna Wiggins
Using GIS to Predict Erosion Hazard Along Lake Superior

Carol A. Johnston and James Salés

Shoreline living is as appealing along the Great Lakes as it is on any coast, providing spectacular vistas of the ever-changing lakes. The changing face of the shoreline itself, however, can threaten buildings and roads built precariously close to the edge (Figure 1). Shoreline erosion is difficult to prevent, especially along a lake as powerful as Lake Superior or the other Great Lakes. A better approach is to understand where and how fast shoreline erosion is likely to occur, and protectively zone erosion-hazard areas to prevent unwise development too close to the lakeshore. Geographic information systems (GIS) are becoming an important tool for identifying and managing these shoreline erosion hazard areas.

Coastal erosion along the Great Lakes has been documented along much of the coast line, but studies (see References) were lacking for the north shore of Lake Superior, due in part to its sparse development. The coastal road connecting Duluth, Minnesota and Thunder Bay, Ontario was not completed until the 1920s, and there are few population centers along the way. However, its relative proximity to the thriving Minneapolis/St. Paul urban area has made the Minnesota north shore increasingly popular for resort and second-home development. With change on the horizon, pro-active planning now can help avoid economic losses in the future.

Beyond the need to determine past shoreline recession rates is the need to predict the location and extent of future erosion. Such coastal characteristics as geology, slope geometry, and groundwater seepage are related to erosion rate (see References). However, most studies of these causal factors rely on data from relatively short time intervals, several years or less. Short study periods may exclude infrequent lake events that are critical to long-term erosion rates, such as major storms and high lake levels.

The purpose of this study was to use GIS to predict and depict recession of the Minnesota Lake Superior coast based on coastal characteristics and historical recession rates. A predictive approach was used in lieu of a comprehensive survey because:

1) There were few fixed reference points suitable for accurate measurement of long-term shoreline recession.
2) A predictive model enabled information from sites with good reference points to be extrapolated to those without.
3) Relating shoreline recession to coastal characteristics identified factors with the greatest influence on long-term shoreline erosion, which could help coastal engineers develop more effective remedial actions for shore protection, and
4) Money for acquiring new coastal zone information was limited.

Coastal Characteristics

Minnesota's Lake Superior shoreline extends 314 km from Wisconsin to Canada, facing southeasterly (Figure 2). Its largest urban area, Duluth (population 85,400), lies at its southern terminus. Other cities along the coast (Two Harbors, Silver Bay, Grand Marais) have populations of less than 4,000.

Existing GIS databases and maps provide some information about land use, geology, and topography of the Minnesota Lake Superior shoreline (Table 1). Most of these data sources were prepared during the late 1970s under Minnesota's former Coastal Zone Management Program. Coastal land use and soils data had been digitized using EPPL6, an early raster GIS developed for the Minnesota Land Management Information System (MLMIS), and were converted for this study to a format compatible with EPPL7, its PC-based successor. Original copies of 1:24,000 surficial geology maps were digitized under this study using PC-ARC/INFO.

Like the rest of the Lake Superior drainage basin, the Minnesota coastal zone is predominantly forested (91 percent; Table 2). Year-round and seasonal residential constitute 3 percent of the coastal zone.
whereas agriculture constitutes only 2 percent. Although 43 percent of the drainage basin is in public ownership, only 4 percent of the coastal zone is public land, mostly in state parks. Commercial and manufacturing acreage is minor (0.4 percent).

The Lake Superior drainage basin contains both the highest (2301 feet) and lowest elevations (602 feet) in Minnesota (Figure 2). The area is underlain by Precambrian bedrock, and shallow to bedrock soils dominate the shoreline from Silver Bay to Grand Marais. Clayey soils are common from Duluth to Silver Bay, whereas sand and gravel deposits occur primarily north of Grand Marais. Surficial geology of the coastal zone is similarly distributed among three main groups: bedrock exposures (42 percent), clay & silt deposits (31 percent), and sand & gravel deposits (20 percent).

Shoreline Recession

Aerial photo measurements of shoreline recession were made at sites selected to represent different geologic types (Figure 2). Because the rate and variability of erosion would be highest for shores composed of clay or sand and gravel deposits, these areas were sampled more intensively than were bedrock shores. Shore location was determined using current and historical aerial photographs (Table 3). Long-term (35- to 41-year) shoreline recession rates were measured by comparing shoreline location on the earliest available aerial photos (1930s) with shoreline location at the same point on the 1975 photographs. A Zoom Transferscope was used to mark all reference points (houses, road intersections) common to both dates of photos, and at least three points per photo which could be identified on 1:24,000 U.S. Geological Survey maps to determine photo scale. Measurements were made using an etched glass
### TABLE 1. GIS data files used and created by this project. USGS = U.S. Geological Survey.

<table>
<thead>
<tr>
<th>Database</th>
<th>Use</th>
<th>Data Structure</th>
<th>Scale or Resolution</th>
<th>Source of Digital Database</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erosion measurements</td>
<td>final erosion maps</td>
<td>point</td>
<td>1:24,000</td>
<td>this study</td>
<td>Johnston et al. 1994</td>
</tr>
<tr>
<td>Shoreline, erosion hazard</td>
<td>final erosion maps</td>
<td>line</td>
<td>1:24,000</td>
<td>this study</td>
<td>Johnston et al. 1994</td>
</tr>
<tr>
<td>Cultural features</td>
<td>final erosion maps</td>
<td>vector (lines and polygons)</td>
<td>1:100,000</td>
<td>USGS Digital Line Graph</td>
<td>Johnston et al. 1994</td>
</tr>
<tr>
<td>Surficial geology</td>
<td>predict erosion hazard for final maps</td>
<td>vector (polygons)</td>
<td>1:24,000</td>
<td>this study</td>
<td>Green et al. 1977</td>
</tr>
<tr>
<td>Land use</td>
<td>background information</td>
<td>raster</td>
<td>2.5 acre cells</td>
<td>Land Management Information Center</td>
<td>Minn. State Planning Agency 1978</td>
</tr>
<tr>
<td>Soils</td>
<td>background information</td>
<td>raster</td>
<td>2.5 acre cells</td>
<td>Land Management Information Center</td>
<td>Minn. State Planning Agency 1978</td>
</tr>
<tr>
<td>Topography</td>
<td>background information</td>
<td>raster</td>
<td>40 acre cells</td>
<td>Minnesota Land Management Information System (MLMIS.00)</td>
<td>—</td>
</tr>
</tbody>
</table>

### TABLE 2. Land use of the Minnesota Lake Superior coastal zone. Data derived from GIS summary of coastal zone land use database.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>Area (ha)</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOREST &amp; SILVICULTURE</td>
<td>77493</td>
<td>90.7%</td>
</tr>
<tr>
<td>AGRICULTURE</td>
<td>1773</td>
<td>2.1%</td>
</tr>
<tr>
<td>COMMERCIAL &amp; MANUFACTURING:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mineral extraction, active</td>
<td>20</td>
<td>0.0%</td>
</tr>
<tr>
<td>Mineral extraction, inactive</td>
<td>38</td>
<td>0.0%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>11</td>
<td>0.0%</td>
</tr>
<tr>
<td>Commercial, retail</td>
<td>80</td>
<td>0.1%</td>
</tr>
<tr>
<td>Commercial, service</td>
<td>23</td>
<td>0.0%</td>
</tr>
<tr>
<td>Dwelling w/ attached commercial</td>
<td>130</td>
<td>0.2%</td>
</tr>
<tr>
<td><strong>SUBTOTAL</strong></td>
<td>302</td>
<td>0.4%</td>
</tr>
<tr>
<td>RESIDENTIAL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residential, year round</td>
<td>1651</td>
<td>1.9%</td>
</tr>
<tr>
<td>Residential, seasonal</td>
<td>683</td>
<td>0.8%</td>
</tr>
<tr>
<td>Mobile homes, year round</td>
<td>101</td>
<td>0.1%</td>
</tr>
<tr>
<td>Mobile homes, seasonal</td>
<td>30</td>
<td>0.0%</td>
</tr>
<tr>
<td>Abandoned buildings</td>
<td>58</td>
<td>0.1%</td>
</tr>
<tr>
<td><strong>SUBTOTAL</strong></td>
<td>2523</td>
<td>3.0%</td>
</tr>
<tr>
<td>PUBLIC &amp; RECREATIONAL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public land</td>
<td>33</td>
<td>0.0%</td>
</tr>
<tr>
<td>Semi-public land</td>
<td>13</td>
<td>0.0%</td>
</tr>
<tr>
<td>Wayside</td>
<td>25</td>
<td>0.0%</td>
</tr>
<tr>
<td>Campground</td>
<td>13</td>
<td>0.0%</td>
</tr>
<tr>
<td>Picnic ground</td>
<td>1</td>
<td>0.0%</td>
</tr>
<tr>
<td>Public access</td>
<td>7</td>
<td>0.0%</td>
</tr>
<tr>
<td>State Park</td>
<td>3283</td>
<td>3.8%</td>
</tr>
<tr>
<td>Camping trailer facility</td>
<td>2</td>
<td>0.0%</td>
</tr>
<tr>
<td>Boat docking area</td>
<td>2</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>SUBTOTAL</strong></td>
<td>3379</td>
<td>4.0%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>89470</td>
<td></td>
</tr>
</tbody>
</table>

scale with 0.1 mm increments and 8× movable magnifying lenses, with the aid of a stereoscope to help determine the bluff edge, to which the measurements were made. The distance over which the shoreline receded was divided by the number of years between air photo dates to compute annual erosion rates. To determine if the record-high lake levels of the mid-1980s accelerated shoreline erosion rates, an air photo flight was commissioned in 1988 to provide data on recent shoreline erosion (Table 3). The same process described above was used to measure shoreline recession that occurred between the 1975 and 1988–89 air photo flights. The recession rates were statistically compared with coastal characteristics, and the presence or absence of bedrock was found to be the best predictor of erosion. Non-bedrock areas (glacial deposits, post-glacial beach deposits, clay bluffs, peat deposits) had the highest erosion rates, even when they overlaid bedrock at the shoreline. Although clay bluffs were expected to have the highest erosion rates, post-glacial sand and gravel beach deposits were equally erosive.
Based on these relationships, maps of shoreline surficial geology published by the Minnesota Geological Survey were used to classify erosion hazard. Sites with any of the following geologic types occurring within 60 meters of the shoreline were classified as "high erosion potential": organic deposits, sand & gravel, clay & silt, unsorted glacial deposits.

### Erosion Hazard Maps

A booklet of 39 page-size 1:24,000 maps was produced to convey the erosion hazard information in a form that could be easily understood. Construction of these maps required combining GIS data layers from several different sources (Table 1): point coverages of erosion measurements from the two different time periods, and a line coverage of predicted shoreline erosion hazard generated by this study, were combined with an existing database showing cultural and locational features to create the final erosion-hazard maps for the Minnesota North Shore from Duluth to the Canadian border. Each map shows about four miles of shoreline, classified into one of the following three erosion potential categories:

- **High**: These shoreline areas have high potential for erosion because they have non-bedrock areas at or near the shoreline. Measured erosion rates during the 1930s to 1975 averaged 0.46 ft/yr in these areas, and were as high as 1.1 ft/yr. Nearly all detectable erosion during the 1975 to 1988/89 period occurred in these areas, with a maximum measured value of 4.5 ft/yr.
- **Low**: These shoreline areas have low potential for erosion because they are predominantly resistant bedrock. Erosion rates during the 1930s to 1975 averaged 0.16 ft/yr in these areas, but measured values were as high as 0.64 ft/yr.
- **Unknown**: These areas are of artificial shoreline, such as industrial harbors, or areas in Duluth for which there were no geologic maps.

Geology was the variable used to predict these erosion-hazard categories, but the digitized maps portrayed surficial geology as polygons covering the entire coastal zone, rather than just the shoreline. Therefore, ARC/INFO GIS was used to convert the surficial geology maps into a linear representation of geology at the land/water interface. In order to reduce database size and minimize processing time, a 1-km buffer of the shoreline was clipped from the larger coastal zone geology database. This smaller database was used to classify shoreline reaches into erosion hazard categories. It was relatively straightforward to classify shoreline segments where erodable deposits (clay, sand and gravel) intersected the shoreline, but more difficult to define areas where erodable deposits overlaid bedrock at the base of the bluff, which were equally hazardous. Such areas were usually drawn on the original source maps as having a narrow band of shoreline bedrock, so the GIS was used to identify coastal segments along which erodable deposits came within 60 meters of the shore.

Actual erosion rates (ft/yr) were shown at the points for which aerial photo measurements were made, distinguished by time period with two different point symbols: open triangles (△) for the 1930s to 1975 time period, and closed (▲) triangles for the 1975 to 1988/89 time period. Recession rates below the detection limits of the measurement method used were coded as “BD.” Values less than 0.3 ft/yr were below detection limits for the 1930s to 1975 time period, and values less than 0.6 ft/yr were below detection for the 1975 to 1988/89 time period.

United States Geological Survey 1:100,000 Digital Line Graphs (DLGs) comprised the most detailed vector database that existed of the Lake Superior shoreline, but the shoreline was too generalized at that scale. Therefore, the shoreline was manually digitized from U.S.G.S. 1:24,000 topographic maps, and combined with the erosion hazard classifications derived from the surficial geology maps to generate erosion hazard segments for the final maps (Figure 3).

Cultura. and locational features, such as streams, roads, railroads, political boundaries, and section lines, were derived from the 1:100,000 DLGs. Over 30 DLGs were imported into ARC/INFO, processed to assign attributes, and edge-matched.

The desired format was 8.5-inch × 11-inch maps, aligned parallel to the shoreline (Figure 3). The shoreline’s southeast-facing orientation presented a challenge, because the maps had to be rotated from the conventional “north is up” orientation. The endpoints of each map were determined by recursively locating points every 6096 meters along the shoreline, the straight-line ground distance covered by each map. In order to orient the shoreline parallel to the long axis of the paper, a short routine was written to compute the arc tangent of the slope between page endpoints in real world
FIGURE 3. Erosion Hazard Map for Grand Marais, Minnesota.

coordinates, and rotate the map layers accordingly. A separate ARC/INFO database was created to keep track of successive map angles and boundary coordinates. This database was later used to automatically position each map's north arrow, and expand the map limits to allow a slight overlap for easy reading across maps.

A second GIS routine was written in Arc Macro Language (AML) to find the endpoints of the final map page lengths along the shore, and determine the UTM coordinates of the map page limits. A third routine was written to combine all of the map layers (erosion measurements, predicted erosion hazard, cultural features), rotate and clip all the layers so the shoreline would be parallel to the main axis of the page, rotate and position the north arrow, and add map details that were common to all of the maps (logo, neat line, legend, scale, and map number). In addition, separate AMLs were written to add information unique to each map, such as place name annotation, township, range and section numbers.

Conclusions
The need to define and identify erosion-hazard areas has been targeted as a key goal by the North Shore Management Board, a joint-powers board consisting of county, city, and township governments to provide environmental protection for and guide the orderly growth of the north shore of Lake Superior (North Shore Management Board 1988). This study provides a basis for that erosion-hazard identification. By representing shoreline erosion as a series of easily interpreted maps, the final product can benefit a variety of users: prospective landowners, realtors, loan officers, zoning boards. Furthermore, the maps have already been used as evidence for insurance claims under the Federal Emergency Management Agency's (FEMA) coastal-erosion program.

GIS was essential to this study in several ways. First, it provided a means for quantifying coastal characteristics potentially related to shoreline erosion, which could be used as input variables for developing erosion-hazard predictions. Second, it provided a means for summarizing a large body of coastal data (~11,000 line segments in the shoreline file alone) so that generalizations could be made about coastal characteristics. Finally, it provided an innovative and cost-effective way to extrapolate a limited number of erosion measurements to the shoreline as a whole, based on coastal characteristics. Even in states having detailed shoreline erosion data, the use of GIS could benefit coastal zone management by compiling complex erosion-hazard information into an attractive and easily understandable map format.

Acknowledgments
John Bonde did the shoreline recession measurements and assisted with the design and execution of erosion hazard maps. Research support from Minnesota Sea Grant (DOC/NOAA/NA46AA-D-0G112 and DOC/NOAA/NA90AA-D-0G149) is gratefully acknowledged. This is contribution Number 117 of the Center for Water and the Environment, 23 of the Natural Resources GIS Laboratory, and Jr 319 of the Minnesota Sea Grant College Program.

Selected References


The Evolution of Michigan's Geospatial Data Infrastructure

Michael N. Beaulac, Shirley Businski and David Forstat

The Michigan Resource Information System (MIRIS), in partnership with both the public and private sector, is producing a growing library of digital spatial information. The success of this 14-year effort is due in part to a market-oriented, user-driven approach toward cooperative data development. The accumulating data are generated by Michigan GIS data users whose needs parallel those described by the Federal Geographic Data Committee and within the framework of the National Spatial Data Infrastructure.

Program Beginnings

In 1979, the Michigan Resource Inventory Act generated a unique state-local program that continues to evolve in ways unforeseen at its inception. The act charges the Michigan Department of Natural Resources (MDNR) with developing current use (maps depicting land cover/use from color infrared photography) and a land resource inventory (maps that depict resources, unique areas, areas hazardous to development, etc.) for application by local and state governments. United States Geological Survey (USGS) topographic maps (1:24,000) are the primary source for the digital base map of the state, which includes all the standard map features minus the contour lines.

Supported by state appropriations and cost recovery/user fees, the Michigan Resource Information System (MIRIS) jointly manages, or coordinates with, the Statewide Groundwater Database (SGDB) and the Great Lakes Information System (GLIS), respectively. The Great Lakes Information System, a GIS for the Great Lakes and connecting channels, is used extensively by the MDNR for storage and analysis of resource inventory information. Such as high-risk erosion shorelines, fish spawning, bathymetry, sediments, sand dunes and other environmental features requiring protection or permitting (see Figure 1). The computer production facilities of MIRIS, GLIS and the SGDB serve as an extensive land and water resources information base for the state, local and federal government.

These production facilities, referred to collectively as MIRIS for brevity, also encode spatial data elements into a MIRIS-compatible form on either a cooperative or contractual basis with other public agencies. (Beaulac 1992) The present MIRIS goals include the support of any agency’s interest in developing or accessing digital information to meet their needs. (Cooperating state agencies and the developing information base is summarized in Figure 2)

Regardless of who does the digital encoding, the MIRIS philosophy continues to be one of data ownership and management by that program, agency, or organization in control or responsible for the data source. For example, it is the joint responsibility of the State Geological Survey Division and the local health departments to build and maintain the water-well records in the Statewide Groundwater Database.

A Market-Driven Approach

Any participatory data development/management process requires a cooperative working relationship between the service organization and its “clients.” The MIRIS data-development philosophy is loosely based, in part, on the adoption of many marketing, funding and management strategies used by non-profit organizations. The basic market-driven, management approach adopted by MIRIS:

1) Understand the market and examine the needs of the clients.
2) Segment the market into manageable groups according to the type of product or service needed.
3) Target the main clients to be served (especially targets of opportunity).
4) Develop products, policies, and a strategy to satisfy targeted clients.
5) Communicate the products, services and philosophies to the clients and
6) Deliver on all promises.

(Kotler 1982; Drucker 1990; Lauffer 1984)

The MIRIS marketing-management approach begins with customers or consumers we want to serve well (i.e., GIS users), versus

Michael Beaulac is manager of the Michigan Resource Information System, Michigan Department of Natural Resources.

Shirley Businski and David Forstat co-manage the Statewide Groundwater Database, Michigan Department of Natural Resources.
selling, which starts with a set of products to be pushed into available markets. (Kotler 1982) Further discussion of how this process operates in Michigan is described in Beaulac (1993).

Organizing GIS Users

Successful public sector, GIS-oriented marketing does have its downside. The steady accumulation of digital spatial information creates additional needs related to such issues as on-line data access, data standards and organization of users into manageable networks. This parallels issues addressed by the National Spatial Data Infrastructure (NSDI) and the Federal Geographic Data Committee (FGDC). Both entities emphasize how “geospatial” data are organized and how users coordinate “to facilitate and foster partnerships and alliances within and among various sectors” (FGDC 1993; MSC 1993).

The FGDC recognizes the need for developing standards, a framework and themes for basic geospatial data, and creating a distributed clearinghouse for finding and accessing geospatial data. All require the cooperation of the non-federal sectors. The Mapping Science Committee (1993) further notes that the “infrastructure” portion of the NSDI should include, not only the materials and technology, but people necessary to acquire, process, store and distribute such information to meet a variety of needs.

The success of the NSDI anticipates a workable organizational structure of GIS users within each state. With an expanding GIS, MIRIS is focusing its efforts on organizing the various GIS users into more manageable groups. This “segmentation” approach results from the observed differing needs of each defined group, and the perception that MIRIS and others can better service these needs by addressing them within a particular group or network.

The organization of GIS users or networks takes many approaches in Michigan, as elsewhere. Here are three organizational groups:

1) **Thematic:** groundwater database users and networks
2) **Governmental:** Department of Natural Resources, and state government users and associated networks
3) **Statewide:** general GIS users organized within the experimental three-year, statewide IMAGIN project

Each of these developing networks poses a unique challenge to MIRIS and the GIS community in terms of coordination and cooperation.

**Groundwater: A Specialized User Network**

Initiatives to develop a statewide groundwater database began in the early 1980s in response to the needs expressed by local public health offi-
s and recommendations of the Governor's Cabinet Council on Environmental Protection. Initial efforts of local and district health departments, under the guidance of Western Michigan University's Science for Citizens Center, were supplemented by state-level activities. The Statewide Groundwater Database (SGDB) is now an ongoing, jointly managed program between MIRIS and Geological Survey Division (GSD), and currently contains over 100,000 geographically referenced water-well records covering most of Michigan's 83 counties.

(Figure 3, Businski and Beaulac 1993)

Computer software developments have aided the growth of the SGDB. MDNR developed the WELLKEY data management program for data entry, query and printing of well record data. Michigan State University's Center for Remote Sensing developed the C-MAP program for digitizing, computer mapping and initial GIS analysis of the MIRIS, SGDB and other data layers. C-MAP allows for the transfer of data to and from other GIS or non-GIS software programs, such as ARC/INFO, ERDAS, SURFER, GEBASE and EPA's STORET database.

Many local and regional agencies and universities are involved in the SGDB either through a cooperative effort with the MDNR or participation in a Groundwater Education in Michigan (GEM) project. (GEM is a statewide program funded by the WK Kellogg Foundation.) These data users are developing GIS applications that have groundwater as their focal point. One significant application is wellhead protection, undertaken cooperatively with US Environmental Protection Agency, state and local agencies.

Some of these local GIS users have organized regional groundwater user networks for themselves. Groundwater-oriented, C-MAP user groups now exist in southeast and southwest Michigan with others starting elsewhere. These groups, comprised primarily of local and regional health and planning department staff, meet periodically for hands-on demonstrations, discussions and general sharing of ideas and technologies related to GIS groundwater applications.

The WK Kellogg Foundation, recognizing the need for a coordinated effort in the area of groundwater database development and application, recently awarded a two-year grant to the MDNR to develop a formal, statewide, groundwater database user network. The objective of this user-driven, state-local partnership is to provide network members with training and education in the verification and automation of water-well and groundwater database applications, and to support a forum for discussion of groundwater-related GIS issues.

**Bridging State Government Networks**

The number of GIS users within the MDNR, outside the MIRIS/GLIS/SGDB organization, has signifi-
cantly increased during the last five years. The Divisions of Land and Water Management, Geological Survey, and Wildlife initiated, and continue to dominate, the technology within the department. Other divisions, either through direct contract or in cooperation with MIRIS, are in the initial phases of data-layer development.

MIRIS staff organized and chairs an ad hoc, department-wide GIS users group. Nearly all natural resource-oriented (non-administrative) divisions are represented. Meetings are every two-to-three months, or based on need. User group issues include data and other technical standards, data sharing, GPS, cooperation on new data-layer development, training and workshops.

The department’s GIS user network is composed of representatives from those divisions presented in Figure 2. Most, but not all user groups represented, have direct access to the MIRIS server for spatial data needs. Those without a server connection receive digital files through some other electronic media. Similarly, several other departments within state government, such as transportation, public health and Office of Management and Budget, have specialized GIS data layers that are “shareable.”

The State’s relationship with the federal government relative to GIS is beginning to increase. The encoding of county soil survey information has occurred in cooperation with the USDA Soil Conservation Service since the mid-1980s. Other data layers, such as floodplain maps, watershed boundaries, and various hydrologic studies, are developed as federal agencies provide funds. These data layers and federal agencies are described in Table 1.

**TABLE 1. Cooperating Federal Agencies for GIS Data/Projects**

<table>
<thead>
<tr>
<th>Federal Agency</th>
<th>Data Layer/Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Agriculture</td>
<td>soils</td>
</tr>
<tr>
<td>Army Corps of Engineers</td>
<td>watershed mapping</td>
</tr>
<tr>
<td>Environmental Protection Agency</td>
<td>coastal zone</td>
</tr>
<tr>
<td></td>
<td>water well mapping</td>
</tr>
<tr>
<td></td>
<td>wellhead protection</td>
</tr>
<tr>
<td></td>
<td>STORET data base</td>
</tr>
<tr>
<td>FEMA</td>
<td>floodplain maps</td>
</tr>
<tr>
<td>US Geological Survey</td>
<td>watershed mapping</td>
</tr>
<tr>
<td></td>
<td>hydrologic studies</td>
</tr>
</tbody>
</table>

**IMAGIN: A Statewide Network**

A more efficient alternative to data development, distribution and partnership with all levels of the public sector, however, may ultimately result from a three-year experiment called the Improving Michigan Access to Geographic Information Network (IMAGIN) Project.

The IMAGIN Project is a unique multi-agency effort funded by the WK Kellogg Foundation to help state government respond to the growing need for digital geographic
information, training and data exchange. It is expected to help coordinate GIS users, and respond to three primary GIS needs: 1) lack of trained staff, 2) cost, and 3) unavailability of data. The IMAGIN Consortium includes the Michigan Department of Natural Resources—MIRIS, Michigan State University Center for Remote Sensing, the Library of Michigan and the Legislative Service Bureau.

Consortium Structure and Function

The project includes a statewide IMAGIN Data Sharing Network (IDSN), a central data archive, and a GIS training facility available to public agencies. The network is designed to improve MIRIS data access and distribution, enhance data exchange efforts, and assure a continuous archival record of geographic and related data in Michigan.

The Library of Michigan supports the central data archive and is the distribution site for all network source geographic information conforming to agreed-upon standards. The Center for Remote Sensing serves as both a GIS-training and education facility, and a quality-control center for GIS data supplied by network members. The quality-control center checks incoming digital information prior to its acceptance into the archive and distribution back to users.

The IMAGIN network (Figure 4) is becoming a focal point for public GIS users and allows for the exchange of quality-controlled, non-proprietary GIS and related data sets created by other network members. Presently, membership is limited to:

- Local units of government
- Public colleges and universities
- State departments and legislative agencies
- Regional planning commissions or agencies
- Michigan-specific units of federal agencies
- Non-profit public organizations approved by the IDSN Board

IDSN Board Composition and Responsibilities

The four consortium members established bylaws to govern a board and guide the network agreement process and members. The Board responsibilities allow for the appointment of committees and the delegation of duties. The 13-member board composition includes a cross-section of both consortium and non-consortium members including:

<table>
<thead>
<tr>
<th>Consortium Members</th>
<th>Non-Consortium Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MI Dept. Natural Resources</td>
<td>2 Local units of government</td>
</tr>
<tr>
<td>1 Legislative Service Bureau</td>
<td>1 Federal agency</td>
</tr>
<tr>
<td>1 Library of Michigan</td>
<td>2 Public colleges/univ.</td>
</tr>
<tr>
<td>1 Center for Remote Sensing—MSU</td>
<td>1 Regional planning agency</td>
</tr>
<tr>
<td></td>
<td>2 State agencies</td>
</tr>
<tr>
<td></td>
<td>1 Other non-consortium member</td>
</tr>
</tbody>
</table>

The IDSN Agreement

The IDSN agreement, sent to and signed by prospective IDSN members, contains several definitions, data-use restrictions/ownership limitations, protocols for data quality, source, format, transfer and other issues. A key component of the agreement is the required quality and source reporting, also known as the data documentation or dictionary.

DATALOGR, a data-documentation program, is available in digital form and accompanies all data sets delivered to or from the central archive. The DATALOGR program requirements and fields are loosely based upon the federal metadata standards.

While data sharing in Michigan currently occurs both informally and through the agreement process described above, there are few examples where it is formally institutionalized. (Brusegard and de Last 1992; Carter 1992; Dangermond...
The IMAGIN Project, if successful, is expected to help institutionalize this process and serve as a model for statewide coordination of GIS beyond the environmental and health professions.

Challenging Times

Michigan is faced with having to make do with available resources to manage the growing demand for GIS statewide. One advantage has been a GIS history longer than many other states and a user-driven, marketing approach to management and data layer development. This has allowed users of MIRIS to gradually foster trust in its GIS products and services and establish a growing advocacy group supporting the program. Their voice has helped keep MIRIS funding alive during the present budget-cutting, government-downsizing period.

New and innovative approaches to develop and maintain a network of growing GIS users and to provide better access to digital information is proving to be a challenge. And even greater demands are soon expected from the FGDC and the IDSN. Greater incentives to promote uniform, statewide data standards, data sharing and a more formalized network, such as a new legislatively created state GIS board or office, will help encourage cooperation and truly make state government more efficient.

References


The Ways and Means of Redistricting in Wisconsin

Ken Strasma and Brenda Haskins

Editor's Note: A map illustrating the redistricting results in Wisconsin appears in this issue's Feature Map on p. 73.

According to Wisconsin's state Constitution, the Wisconsin Legislature is required to reapportion the state Senate and Assembly during the first session after the U.S. Department of Commerce conducts a decennial census. The census data as of March 1991 showed the population of Wisconsin had shifted and that the 99 assembly districts and 33 senate districts were clearly malapportioned and would have to be redistricted to approximate the U.S. Supreme Court designated target of "one person, one vote."

The state's population had grown to 4,891,769, which meant that each of the 99 assembly districts should have as close to 49,412 as was practicable. The population of one district was 65,494—32.8 percent over the new ideal population. Another was 43,641, or 11.66 percent under the new ideal population. While no legislator wanted their district to change, it was inevitable.

Wisconsin's Political Environment

In 1991, the Assembly and State Senate were both controlled by a majority of Democrats; the governor was a Republican. The Legislature had begun holding public hearings in the fall of 1991 to solicit public input on the principles of redistricting including population equality, voting rights, community of interest, compactness and contiguity, political fairness and voter disenfranchisement. By January 31, 1992, the Senate and Assembly Democrats had each introduced and held hearings on their plans. We had expected the Republicans to introduce alternative plans; however we were surprised by their decision to go directly to court. Some political observers speculated that the Republican strategy was to bypass the legislative process, and instead to gamble that they could do better in federal court.

Any legislator had the legal right to introduce a redistricting plan during the 1991-92 legislative session. However, the expense and complexity of the computer equipment used for redistricting made it impractical for each legislative office to have its own redistricting operation. Instead, each of the four legislative caucuses (Assembly Democratic, Senate Democratic, Assembly Republican and Senate Republican) had computer equipment and staff to assist their members in drawing redistricting plans. In addition, the legislature provided a public-access terminal and support staff to any group or individual who did not want to work through a legislative caucus.

As the Assembly Democratic Caucus Redistricting Team, we had two goals: 1) To assist our members in drawing the best possible redistricting plan, and 2) to defend it against the inevitable challenges in the legislative process and in the courts. Because there is no agreement on what would constitute the "best" redistricting plan, we had to anticipate and be prepared to counter any possible attack to the Assembly's plan.

Equal population, protection of minority voting rights, and avoidance of "intended political consequences" were all criteria that had been used in defending and attacking redistricting plans over the last decade. In the 1990 redistricting, those basic criteria blossomed into hundreds of statistics that different groups used to attack and defend plans. As we worked with legislators, community groups and interested individuals on drawing draft plans, they suggested new statistical measures that we then incorporated into our system.

Ensuring Political Fairness

To illustrate the political fairness of our proposed plan and to measure the political fairness of others' plans we used an index we called ADC1 (Assembly Democratic Caucus Index #1). The ADC1 is an average of the Democratic performance in the following 11 state-wide races: governor 1982; president and U.S. Senate 1984; governor, attorney general and state treasurer 1986; president and U.S. Senate 1988; and governor.
attorney general and state treasurer 1990. The ADC1 is generated by taking the total Democratic vote cast in these 11 races divided by the total vote (Democratic and Republican) for the 11 races. We chose this methodology over the alternative of averaging the 11 individual percents Democratic so as to emphasize races with higher turnout. The reason? We anticipated an extremely high voter turnout in November 1992, the first election to be held under the new redistricting plan.

The ADC1 includes all 11 of the statewide races for which we had election results on the redistricting computers. Both parties had the opportunity to request that races be added throughout the process, and the races for attorney general and state treasurer were added at the request of the Republicans.

Our goal in developing ADC1 was to create an index that would accurately reflect the partisan impact of any given redistricting plan. Wisconsin is a competitive state, where races are often decided by factors unique to that election rather than straight partisan allegiance. The independence of our electorate is illustrated by the fact that, at the time, we had a Republican governor, Democratic majorities in the state Assembly and Senate, a Republican and Democratic U.S. Senator, four Democrats and five Republicans in the House of Representatives. Knowing that any set of election results would bring its own circumstances that might not be applicable in predicting the likely outcome of legislative races, we categorized the various kinds of races and circumstances, and worked to get a broad distribution among the various categories.

The Tools

The basic tool for redistricting was a program called Wiscore developed by a nonpartisan development team. Wiscore was shared among the four legislative caucuses, and the Department of Administration representing the Republican governor. The program was written in ARC/INFO and ran on IBM RS-6000 workstations.

With Wiscore, all caucuses had the ability to build redistricting plans by assigning counties, minor civil divisions, wards or census blocks to districts. At the completion of each assignment, Wiscore would provide basic statistics, including district population and racial breakdown. In a process as rapidly changing and deeply politicized as redistricting, we knew that our plans would be subject to scrutiny and attack on a constantly changing set of criteria. To draw the best possible plan and to defend it, we would need the ability to run reports on the fly, and to add new statistics at a moment’s notice. For example, we were asked at one point to generate a listing of Democratic wins and losses in a proposed district. The flexibility of our setup allowed us to comply with this and other unanticipated requests throughout the process.

As useful as Wiscore was, for political and technical reasons we decided against working with the Wiscore team to enhance the program package to meet our specific needs. Politically, we were cautious in working with the Wiscore development team, because as a non-partisian entity, they had a stated policy of sharing any changes to Wiscore with all four caucuses. So any innovation that we added might end up helping our opponents.

Aside from the political and security considerations, there were technical reasons why we chose not to work on enhancing Wiscore:

- The package was being developed throughout the process, so any "hard-coded" changes to the Wiscore program code would have to be redone for each new release of the Wiscore package.
- INFO the database portion of the ARC/INFO language in which Wiscore was written, did not have the capabilities we needed.

Our solution was a primitive but very effective form of distributed processing using an IBM PC linked to the RS 6000 workstation. Reports could be run on the PC while plans were being updated on the workstation without slowing Wiscore. Once the link into Wiscore was programmed, we could modify reports on the PC without worrying about Wiscore.

We maintained a dBase IV file on the PC that contained names of all of our redistricting plans. To run reports, the user would select the name of the plan file from a menu on the PC. A dBase program would then write an AIX shell program that would call C and SAS programs to read and translate the desired plan file on the workstation, and a PCplus script file to automate communication between the PC and the workstation.

Once dBase had written the custom PCplus and AIX programs for the plan in question, it would invoke PCplus on the PC. The PCplus script would then transfer the custom AIX program to the RS6000 and invoke it on the workstation. The AIX program would first call C program to translate the internal INFO file (the native file format of the Wiscore package) to an ASCII file. INFO was never run: it would have been too slow, and we were worried about unpredictable results if we tried to run it on a file already in use by ARC/INFO in Wiscore.

Once the C program had translated the data from INFO’s internal format to an ASCII file, SAS would be run to link block or VTD (Voting Tabulation District) codes into the file. Wiscore plan files are simply a series of district numbers that correspond to county, minor civil division, VTD or census block in separate files depending on the plan’s...
level of geography. In theory, we could have transferred just the plan file and linked it to geography on the PC, but we were concerned that any undocumented change in the order of plan file elements, or the addition or subtraction of a geographical unit would invalidate the ordered relate. Our response to this concern was to link in unique geographic identifiers on the workstation, and from the same files used by Wiscore.

As soon as the SAS session had finished writing the final linked file, AIX would initiate transfer of the file back to the PC. Once loaded on the PC, the user could choose from a menu of more than 20 different reports. The whole process would take a little more than 1 minute, which was at least 100 times faster than running custom reports in SAS.

Wiscore noticeably. So we usually relied on the PC for statistical work.

This distributed processing solution was developed and used throughout the plan-drawing process.

A Strong Case

As technical staff, it was not our role to decide what statistics were important, but rather to provide whatever numbers were requested. (Occasionally, we did have to gently dissuade people from using some creative but statistically invalid measures. One of our favorites was the person who added together the percent Democratic for the three Assembly districts that made up a particular Senate district to get that district’s percent Democratic. He was puzzled when the resulting number was greater than 100 percent.)

Because we were pursuing the legislative process, we had to get a majority of votes in favor of a plan in both the Assembly and the Senate. One measure we found useful to show legislators was the percent of a proposed district that had been in their old legislative district. We also showed them who had represented the new parts of their proposed district. No legislator wanted their district to change. In fact, one legislator, when told that his district would have to change because the state population had changed, was convinced that we had removed people from his district and wanted us to put them back.

The Democratic plan had evolved slowly through a series of public hearings and open debate in the legislature. The Republicans, for whatever reason, chose not to introduce a plan through the legislative process, thereby avoiding the public scrutiny to which our plan had been subjected. By the time the legislative redistricting plan had passed, we had built up a large inventory of reports that could be run on any plan. This proved useful when the Republican plan was finally unveiled in federal court in April. In minutes, we were able to run the reports we had developed over the last six months.

The federal court trial lasted three days, during which we ran reports and analyses on the Democratic and other plans around the clock. Despite the short notice caused by the last-minute introduction of the Republican plan, the system we had developed and improved throughout the redistricting process allowed us to provide the Democratic legal team with considerable ammunition to counter the shortcomings of the Republican plan. Their plan had high voter disenfranchisement caused by moving many voters into new districts. According to their plan, four Democratic senators, including the majority and assistant majority leaders, who represent even-numbered senate districts were put into odd-numbered districts. Because only even-numbered senate districts were up for reelection in 1952, the four senators would be prevented from running for reelection. Or, they would have been forced out of their districts, pitting them against incumbent senators in another district.

The court agreed with our assessment of the Republican plan, rejecting it as the “more partisan” and “designed to decapitate the Democratic leadership in the Senate.” In the end, the court enacted a hybrid plan that closely resembled the Wisconsin is a very competitive state, where races are often decided by factors unique to that election rather than straight partisan allegiance.
plan submitted by the Assembly Democrats.

Although the politics of redistricting remains as intense and byzantine as ever, the computer technology used in 1992 allowed more public input, scrutiny and empirical analysis than ever before. This is one case where computer technology has helped to move an important public-policy question out of the smoke-filled rooms, and into the light of day.
Assembly Democratic Caucus 1

Index of Partisan Performance in Milwaukee County

Republican
- < 35.0
- 35.0 - 39.9
- 40.0 - 44.9
- 45.0 - 49.9
- 50.0 - 54.9
- 55.0 - 59.9
- 60.0 - 64.9
- >= 65.0

Democratic

3 miles
The Rising Tide of GIS in Minnesota

William J. Craig

Minnesota has been a world leader in the development and utilization of GIS technology, primarily because of a single organization—the Land Management Information Center (LMIC) in state government. LMIC is bigger than ever, with an annual budget of $2 million and a staff equivalent of 32 full-time people. But its efforts are being overwhelmed by the rapidly rising tide of other users in state and local government. This paper is about the nature of that rising tide and the problems it causes for coordination.

In 1977, when LMIC began its operational stage in state government, it accounted for all the state’s GIS activity. By 1982, it had been joined by the Minnesota Department of Transportation (MnDOT), the city of Minneapolis and Hennepin County. Growth since that time has been accelerating (Figure 1). Things are happening so fast that no one knows the real level of activity in Minnesota; the data presented here are a first attempt to provide a current snapshot of GIS activity.

The Rising Tide of State and Local GIS

In 1988, 20 state departments and local government units were using GIS. The level of GIS activity in Minnesota’s state and local governments has more than tripled since then. (Figure 1) In 1992, there were 71 active users and many more who said they were planning to incorporate GIS within the next few years.

LMIC, the Minnesota Department of Transportation (MnDOT), and the Department of Natural Resources (DNR) led the list of six departments at the state level that were using GIS in 1988. By 1992, the number of departments operating GIS had increased to 16. Besides LMIC, MnDOT and DNR, the 1992 list includes:

- Board of Water and Soil Resources
- Education
- Employee Relations
- Health Department
- Housing Finance Agency
- Legislative GIS Office
- Metropolitan Council
- Metropolitan Waste Control Commission
- Military Affairs
- Pollution Control Agency
- Public Safety
- State Fire Marshall
- State Lottery

State government in Minnesota got an early start, but the real growth since 1988 has been in county government. Since that time the number of counties using GIS technology has increased from seven to 35. Figure 2 identifies the counties using GIS technology in 1992. It also identifies another 17 counties that plan to acquire the technology by the end of 1995. Counties use GIS in different ways. Ten counties use GIS to map ownership parcels. With a few exceptions (Blue Earth, Olmsted, Otter Tail and Sherburne) these counties are in the Twin Cities metropolitan area. Counties in the forested northeast tend to use GIS for natural resource management, especially forestry. Water quality, planning or community development activities are the major uses in other counties. A number of counties are using GIS in public safety (to map crime incidents, for example). Using GIS and computer-assisted design for highway engineering and facilities management is also relatively strong now, and it is the most important factor attracting those who will be acquiring GIS in the next three years.

Cities have also been major users of GIS, though activities have been restricted to Minnesota’s larger cities. But the size threshold is dropping (Figure 3). In 1982, only Minneapolis had GIS. In 1992, every city with 40,000 or more people was using this technology—primarily for engineering and facilities management, public safety, and planning and community development. Within the next few years, cities of 15,000 will be using it.

How Is GIS Used?

The range of GIS applications is enormous. Some are limited to urban areas, others to rural areas. Some applications involve a major one-time effort, while others involve automating day-to-day tasks. The following examples illustrate each of these types of applications in Minnesota.

---

William J. Craig is assistant director of the Center for Urban and Regional Affairs, University of Minnesota. He is a past president of URISA, and the recipient of the Edgar Horwood Award for distinguished service in 1993.
Urban Issues

Like any big city, Minneapolis is experiencing urban problems. But these problems are not uniform across the city. If intervention is to work well, it must focus on relatively stable neighborhoods as soon as they begin to show signs of decline. GIS technology has been used in Minneapolis to help direct a new intervention program: CARE (Community and Resource Exchange).

The CARE program works with neighborhood residents to identify and clean-up problem properties.

This program has been very successful, beginning its operations in Minneapolis' Jordan Neighborhood in 1990. Jordan was selected by using key indicators of urban decline: an increase in drug-related arrests, a rise in births to single mothers, more problem pregnancies, and a recent decline in the number of owner-occupied homes. Data about these factors are collected on a regular basis by the police department, the health department and the city assessor.

GIS technology was used to transform this address-based data into summary information by neighborhood across the entire city. Jordan was identified as the neighborhood that could best benefit from the new CARE program and results have shown that the decision was appropriate.

Tax Parcels

Every county has the job of keeping track of and parcels for tax purposes. Maps are a useful way of portraying parcels because they can show important aspects like completeness of the database, the parcel size and its location. These characteristics are important for developing and maintaining an equitable tax base. Sometimes this special application of GIS gets its own name—Land Information Systems or LIS.
Developing and maintaining the maps in Ramsey County, where St. Paul is the county seat, would be an impossible job if it had to be done manually. The county contains 160,000 parcels of land and the initial job of combining records and adding geodetic control would have been difficult without GIS technology. Figure 4 is an example from the Ramsey County parcel maps. The state air grounds appear in the northwest quadrant of the map. The map shows roadways, legal parcels (including some that are no longer available for residential use), and the “footprint” of existing structures.

The real problem is that the job of mapping the county is never done. Every day there are new subdivisions, lots are split, and lots are combined. There is an average of 300 changes per month. There are about 350 map sheets for Ramsey County, so an average a printed sheet is good for about one month before it begins to become obsolete. With GIS technology, the base map is electronic, it is always current, and it is immediately accessible to all cities and to all county departments.

Environmental Issues

Acid Rain

Acid rain falls on much of the state, but some areas are more vulnerable than others. The Minnesota Pollution Control Agency contracted with the state’s Land Management Information Center (LMIC) to help identify those vulnerable areas. A model was built that incorporated critical factors like local soil type and vegetative cover. The most vulnerable areas are those with limited calcium in the soil to buffer the acid rain and those with coniferous trees, which add to the acid problem. Figure 5 contains the final results of this GIS analysis.
Groundwater

Another continuing environmental issue is groundwater quality. Until recently data about ground water were collected by five agencies in an uncoordinated fashion. The agencies each collected operational data from their programs, but these data were never integrated and therefore could not be used for summary analyses of current conditions or for modeling the effects of potential changes. Minnesota has recently started the Groundwater Clearinghouse System, and the LMIC is developing a computer system to bring together data from the five agencies. More important, the clearinghouse effort brings these agencies together to work on common issues. This horizontal integration across bureaucratic boundaries is a frequent side-benefit of GIS technology.

Forest Management

The state of Minnesota manages 2.8 million acres of commercial forest land through the Department of Natural Resources' Division of Forestry. Proper management involves many decisions for areas as small as a few acres or as large as a section (640 acres). Not long ago, field offices were swamped with work, yet were unable to use the sophisticated forest inventory data at the GIS centers. The solution turned out to be down-scaling the data so they could be analyzed on microcomputers, which most field offices already have. LMIC created a customized user-interface that turned a generic desktop GIS program into a tool that has the look and feel of a forest-stand management tool. The result saves time in planning field work and alerts managers to the possibility of conflicting strategies affecting any specific stand. Prior to this innovation, foresters had to page through a foot-high stack of
printouts, and even then seldom got
all the relevant information.

Figure 6 shows a management
map in which various stands are
recommended for some kind of
treatment (clearcut and regenerate,
salvage and regenerate, or thin).
These stands will be visited before
action is taken to verify the com-
puter analysis. Four stands are
shown in need of a new ground sur-
vey (reinventory opportunity)
which can probably be done on the
same visit to the area. A large stand
in the southwest corner of the sec-
tion is labeled "sensitive," which
means that a potential management
conflict exists. This might signal the
presence of old-growth trees, a deer
yard or a recreation area. Any har-
vesting in this stand would there-
fore require extra care.

The Dollars Involved

No one knows how much money is
being spent on GIS, but this impor-
tant question can be addressed in
other ways. With or without GIS,
governments and utilities are spend-
ing large amounts of money on data
about land. In 1976, Wisconsin did
the only known comprehensive
study of the cost of maintaining
land records—surveys, logs and
permits, maps and charts, plans and
studies. Assuming that expenditures
in Minnesota are similar to those in
Wisconsin, one can estimate what
Minnesota is actually spending (Fig-
ure 7). By extrapolation we can esti-
mate that Minnesota is spending
over $133 million each year on land
information. Most of the money (51
percent) is spent by local govern-
ment by counties and cities. The
cost to every man, woman and child
of maintaining these data is $39 per
year. GIS technology could provide
a way to handle land information
more efficiently, thereby saving
money.

Supplementing the monies spent
on maintaining geographic data for
day-to-day operations, LCMR (the
Legislative Commission on Min-
nesota Resources) uses money from
the state's cigarette tax and from the
Environmental Trust Fund (part of
the state lottery) to fund projects re-
lated to Minnesota's natural re-
sources. During the 1991-93 bienni-
um, over $8 million was spent on
such GIS-related projects as: acquir-
ing orthophotos of the state ($2.2
million), producing county geologi-
cal atlases ($1.4 million), finishing
county soil surveys ($1.3 million),
computerizing the National Wet-
lands Inventory and DNR's Pro-
tected Waters Inventory ($1.1 mil-
lion), and continuing county
biological surveys ($1 million). GIS-
related projects will receive LCMR
funding in the 1993-95 biennium of
another $2.7 million.

Who's in Charge?

With all this activity, it is important
to provide as much education, direc-
tion, coordination and leadership as
possible. Many individuals and or-
ganizations are doing what they
can, but more could be done. There
are many players, but no single one
with the authority or the resources
to take a leadership role.

- The Land Management Information
  Center (LMIC) has been one of the
  world leaders in the development
  and use of geographic information
  systems. Developed originally at
  the University of Minnesota, it moved
to state government in 1977 and was
  housed in the Minnesota State Plan-
ing Agency. A decade ago LMIC
  had the resources and the will to pro-
  vide leadership. But there has been
  no growth in staff positions to match
  the proliferation of GIS across state
  and local government.

- The Minnesota GIS/LIS Consor-
tium, a volunteer organization that
  began in 1988 when a group of inter-
  ested persons from the state, local
governments and academic institu-
tions met to share information about
  GIS in a round-table discussion. The
  consortium continues to hold these
  forums and, in addition, publishes a
  newsletter twice a year to share infor-
mation more widely. Beginning in
  1990, the consortium has held an an-
  nual conference and workshops open
to anyone interested. Typically the
  conference draws about 500 people
  from around the state. In 1989, the
  consortium was funded by the In-
terTechnologies Group to look at the
  potential for integrating geographic
data in state government. A number
of excellent inventories and recom-
ended standards resulted and topi-
cal committees formed during this
work (water information, forestry,
and land use) continue to meet and share information.

- **The Governor's Council on Geographic Information.** One problem the consortium has is that it lacks official authority. That was one of the reasons for establishing the Governor's Council on Geographic Information, created by Governor Carlson's executive order in September 1991. Its 18 members represent state and local government, academia and the private sector; its purpose is to promote "the efficient and effective use of resources by providing leadership and direction in the development, management and use of geographic information in Minnesota."

The council first met in March 1992 and has committees working on data coordination and standards. A new committee is looking at local government. However, with limited staff support, progress has been slow.

- **The Minnesota Department of Transportation (MnDOT)** has recently been able to accelerate its program of mapping roads and surface waters and will have the state completely mapped within the next two years. This effort will save other agencies and local governments significant amounts of time and money. Second, MnDOT is taking a lead role in coordinating state efforts to establish GPS (global positioning system) base stations that will make it possible for local survey field crews to make optimum use of satellite signals and accurately locate positions on the ground. This is important to GIS work where high accuracy is required for properly registering and combining maps. Third, MnDOT has a cooperative agreement with the National Geodetic Survey to fund a state advisor. The advisor has been exceptionally useful to state and local offices trying to upgrade their survey base. Unfortunately, the state advisor program is in jeopardy because of federal budget cuts.

All of the players, however, tend to overlook local government, where GIS development is most fragmented. County soil and water conservation districts are getting good guidelines and software from the state's Board of Soil and Water Resources, but their activities are relatively narrow in scope. The Inter-governmental Information Systems Advisory Council (IISAC — part of the Minnesota Department of Administration) has a mandate to provide direction, but very little budget. One of the most promising aspects for Minnesota counties is that almost all of the smaller counties (70 out of 80 small counties) are in one of three computing cooperatives. If solid programs and standards could be developed, they would be relatively easy to disseminate. But to date, no state body has the mandate or the resources to help direct local government GIS activities.

**What Should We Do?**

Lots of money is being spent on geographic information, but it's not clear that we are getting maximum return on the taxpayer's investment. Some data are being collected twice by different agencies while other important data items are missing. And much data collected for single purposes prove useless to others because definitions were too narrow. Significant effort should be put forward to coordinate activities and make it easier to share data.

Two reports by the National Research Council (NRC) point toward actions that could be taken by Minnesota and other states. The most recent report is *Toward a Coordinated Spatial Data Infrastructure for the Nation* (1993). While the report focuses on federal efforts, its recommendations are equally valuable for states. It calls for an expansion of both the authority and the effort of the coordinating body. And it calls for improving access to information about spatial data. The Federal Geographic Data Committee (FGDC) has started some of this work and its efforts will help states also.

Minnesota could learn from these federal efforts and: 1) empower the Governor's Council on Geographic Information to develop policies and strategies that encourage data sharing, 2) create or adopt standards, 3) generate an inventory of spatial data and data users, and 4) develop a strategic plan for filling in missing pieces of the state data infrastructure.

A final recommendation is for the state to help meet the GIS needs of local government. The largest need is for technology assistance including training, guidelines, and a hotline to answer questions. The NRC report *Need for a Multipurpose Cadastre* (1980) stated, "[w]e recommend that each state authorize an Office of Land Information Systems..." to provide coordination and guidance. Such support would save much time and money at the local level and deliver a uniform network of systems supplying comparable data across county lines. There is no reason for such an office to limit its services to land ownership; city and counties have many other needs.

We can look next door for a model of how such support for local government might be funded. Wisconsin has developed a Land Information Board that distributes money to counties for upgrading their land information systems, maps and data about tax parcels. Over $6 million is raised each year from fees charged to register deeds. Two-thirds of these funds are automatically returned to counties whose plans to upgrade their land records have been approved by the state board. The remainder is available for special project grants to the counties. The board has both the power and the resources to coordinate the development of county land information systems. This model of tying the source of funds to the group that ultimately benefits is intriguing, and perhaps more can be made of this for funding other GIS needs.
GIS Integrates Water-Quality Data and Remote-Sensed Images

Joel J. Mouradian, Arthur S. Brooks and David W. Bolgrien

The Great Lakes Ecological Process Pilot Project (GLEPP)—a multi-agency effort (see sidebar)—involves compiling a multi-resolution digital database for ecosystem characterization and modeling. This project takes place on Lake Michigan’s Green Bay and surrounding land area. The goal is to combine remote-sensed data and on-site water-quality data in a GIS to help better understand water quality in lake systems.

Data collected on the project include the following:

- daily aircraft multispectral scanner (MSS) images with thermal capability;
- daily water-quality sampling along selected transects;
- shipboard sky/water spectral measurements at station locations;
- global positioning system (GPS) locational data of both the aircraft MSS and shipboard water-quality samples;
- airborne visible and infrared spectrometer (AVIRIS) data.

United States EPA aircraft collected the MSS data and respective GPS data at an altitude of 7300 meters along predefined flightlines (Figure 1). The AVIRIS data were collected by NASA aboard an ER-2 aircraft at an altitude of 20,000 meters. Data collection dates were June 10–16, 1992.

The water-quality data were collected to identify and quantify variables influencing spectral reflectance and thermal emissions in Green Bay. The use of surface reflectance spectra in lakes, as compared to oceans, is complicated by increased optical complexity from sources such as surface runoff and rivers. Sampling focused on identifying and quantifying particulates and dissolved matter, which absorb and scatter light. Integration of these data will attempt to identify systematic changes in concentrations of phytoplankton, detritus, suspended inorganic sediments, and dissolved matter, which influence spectral reflectance and thermal emissions. Linking the range and distribution of these variables collected on ship with the MSS imagery collected by aircraft will help identify relationships between upwelling radiance spectra and optical cross sections of water components. These relationships should promote the use of remotely sensed data from various sensors to delineate changes in water quality in lake systems.

The U.S. EPA is preparing the GIS database of the MSS imagery data. This article describes the processes involved with creating the GIS database of in situ water-quality data of Green Bay which was done at the Center for Great Lakes Studies.

Data Point Layers

Sampling of 21 water-quality variables was conducted aboard two ships, the U.S. EPA Research Vessel (R/V) Lake Guardian, which primarily performed collections in the southern area of Green Bay between Green Bay, Wisconsin north to Sturgeon Bay, Wisconsin, and the University of Wisconsin-Milwaukee R/V Neesky (Figure 2), which performed collections in the northern area of the bay north of Sturgeon Bay (Figure 3).

Data were recorded every five seconds using an instrument package connected to a personal computer on board the ship (Figure 4) and subsequently averaged over one-minute intervals. The summarized data were associated with a time and a geographic location and stored in a GIS. ESRI’s PC ARC/INFO software was used.

GPS data collected on board the ships were postprocessed with Trimble Navigation GPS Professional Pathfinder and U.S. EPA EMSS-LV software. To obtain a more accurate position, GPS data files were differentially corrected using a second GPS receiver located...
at the Green Bay airport. Positions were retained as Universal Transverse Mercator (UTM) zone 16 coordinates defined using the North American Datum, 1927 (NAD-27). The GPS positions of the Lake Guardian were taken at approximately five-second intervals; on the Neeskay they were taken every 15 seconds. The difference was due to the availability of a PC to transfer X,Y coordinates onto ASCII text files on a PC aboard the Lake Guardian when arriving and departing pre-defined stations throughout Green Bay. Whereas, GPS data from the receiving unit aboard the Neeskay was only transferred to ASCII text files at the end of each day’s collections.

Although the GPS readings were provided every five seconds from the R/V Lake Guardian, or 15 seconds from the R/V Neeskay, the data collected were summarized for every minute. Therefore, UTM X,Y coordinates were averaged for each minute. The average X,Y coordinates were obtained by calculating a simple mean of the readings from the second half of the previous minute with the readings from the first half of the current minute.

After calculating the average X,Y coordinates from the GPS unit for each minute, some minutes would not have a corresponding X,Y location. This was due to gaps in the stream of GPS readings over the 10 to 14 hour daily cruises. Occasionally, the gaps in the stream of X,Y coordinates provided by the GPS readings would span several minutes. The missing X,Y coordinates would then be obtained from one of two sources:

1) The captain of each ship maintained a detailed log of the ship’s location throughout each cruise using LORAN-C. The ship’s log contains lat/long positions with the corresponding time that the ship was at a predefined station on Green Bay. Other locations recorded on the log include changes to the ship’s course

FIGURE 2. Research vessel R/V Neeskay, from the Center for Great Lakes Studies of the University of Wisconsin-Milwaukee.
to avoid hazards or other vessels. If the missing X,Y location coincided with one of the minutes recorded on the ship’s log, the lat/long position from the log could be used. A PC ARC/INFO SML macro was written to easily project lat/long coordinates to UTM using PC ARC/INFO’s PROJECT command.

2) Remaining X,Y coordinates were then interpolated. A straight-line method of interpolation was used since the ships usually traveled at a relatively constant speed of about 10 knots. Some arbitrary adjustments were made after the interpolation process to account for movement of the ship at a station due to drifting. A PC ARC/INFO SML macro was written to carry out the interpolation of the missing X,Y coordinates.

Once the X,Y coordinates were averaged or interpolated, they were combined with water-quality data. These data files were contained in Quattro Pro spreadsheet files, which were written to ASCII text files. From there, PC ARC/INFO SML macros were used to convert the files, one for each collection date, into point coverages.

The collection cruise routes could be viewed on a base map layer of bathymetry contours and the coast.

Project Partners

The Center for Great Lakes Studies at the University of Wisconsin-Milwaukee was a partner in the project along with the Lockheed Environmental Systems and Technologies Company contracting through the U.S. Environmental Protection Agency (EPA) Environmental Monitoring Systems Laboratory, Las Vegas (EMSL-LV), and the National Aeronautical and Space Administration (NASA) Ames Research Center. Also cooperating in the project were the U.S. EPA Great Lakes National Program Office in Chicago, the Green Bay Metropolitan Sewage District, the Wisconsin Department of Natural Resources, and the University of Wisconsin-Green Bay.
line of Green Bay. The coast line coverage and a coverage of bathymetry points were provided by EMSL-LV. The bathymetry point coverage was based on data obtained through the National Oceanic and Atmospheric Administration (NOAA). The bathymetry point coverage consisted of a square grid of X,Y,Z points; Z being the depth at location X,Y. The points were located 100 meters from each other with approximately 300,000 points necessary to cover Green Bay. Contour lines at two-meter intervals were drawn from the bathymetry points using Golden Software's Surfer package.

**FIGURE 6.** Example of separate circular buffers used to identify which data points will be used in a weighted average to be assigned to a grid point.

**Grid Point Layers for MSS Images**

The bathymetry point coverage also served as a means to relate the water-quality data to the MSS images. The points in the bathymetry coverage were in a 100-meter square grid. This grid served as a common locational reference for both the water quality data and the MSS images. Each grid point, which had water-quality data assigned to it, represented the center point of a pixel from the remotely sensed images.

To identify which grid points to assign water-quality data, a buffer was produced around the initial data points. The buffer was used to clip points from the total grid of bathymetry points, which were near the data points (Figure 5). The width of the buffer was set as 71 to ensure that all data points would be associated with a grid point (i.e., the distance between grid points on a diagonal from each other is equal to the square root of \((100^2 + 100^2)\), half of that distance is 71).

This subset of bathymetry grid points along a cruise route needed water-quality data associated with each point. To do so, circles were drawn with each grid point as a center point and the radius equal to the width of the initial buffer (Figure 6). The PC ARC/INFO buffer function was not used since two or more overlapping circular buffers of nearby points are dissolved into one buffer. The necessary result was to have one circular buffer for each grid point. Point-in-polygon analysis was done to associate data points within each circular buffer to the corresponding grid point.

The grid points were given the same attributes, i.e., water-quality variables, as the initial data points. However, the values assigned to these attributes were weighted averages with those data points closer to the grid points being given a higher weight in the average. The following inverse distance average for-
The formula was used in assigning data values to attributes for the grid points (from Shepard 1968):

\[ X = \frac{\sum x \cdot d^{-2}}{\sum d^{-2}} \]

This weighted averaging function is commonly used to interpolate sampled data in order to produce surface models.

At the time of this writing, the process of identifying the appropriate grid points and assigning weighted averages is complete. Each data-collection date has a corresponding grid point coverage in the GIS. The data in these coverages are currently being analyzed at the U.S. EPA EMSL-LV to identify relationships between water-quality data and data contained in the MSS images.

Reference
FEATURE MAP

In this issue . . .

The feature map and inset on p. 73-74, created by the Assembly Democratic Caucus of the Wisconsin State Legislature, illustrates a composite view of Democratic Party performance in statewide elections spanning the nine-year period, 1982-1990. The map presents one of the several voting-pattern indexes used by the Democratic caucus as part of the 1991 redistricting process in Wisconsin. The redistricting process that led to the creation and use of this index is further described in the accompanying text and in the Feature article by the same authors on p. 69.

Editorial Intent

Many of us derive great pleasure from viewing maps. Often we feel that the enjoyment maps bring us is something we have felt for a long time, something we grew up with. It may be difficult for us to pinpoint the precise appeal that maps hold, but often they may provide new understandings and stir our imaginations.

Even though well-designed and informative maps have existed for decades, the growth of computer mapping and geographic information systems has allowed us to experiment and develop new and exciting ways of analyzing and displaying data in a map context. 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. Because computer-generated maps have become an important part of URISA’s domain, the URISA Journal showcases the efforts of map-makers by featuring an exceptional map product in each issue. The word “map” is used in a very broad context and can include remotely sensed images or other graphics.

Our intent is to feature a wide variety of maps, hoping to inspire others to learn and apply good ideas and techniques. We look for maps that are easy to read, have a pleasing appearance and clearly communicate the map’s purpose. New methods of analyzing and presenting phenomena through mapping techniques are also desirable, as is presenting a map that is particularly pleasing or attractive.

If you have produced a map or know of one you believe should appear in the URISA Journal, please contact the editors.

Ted W. Koch
Wisconsin Assembly Democratic Caucus Index #1
Ken Strasma and Brenda R. Haskins

Editor's Note: The Feature Map is placed on p. 73 because of the layout of page signatures.

The purpose of Assembly Democratic Caucus Index #1 (ADC1) was to show the voting patterns of the Wisconsin electorate in every town, village and city within the state. The index portrayed on the map is composed of an average of the democratic party performance in 11 statewide elections during the period 1982-1990. ADC1 was used to analyze proposed redistricting plans in order to detect and counter any "intended political consequences" (otherwise known as gerrymandering).

Politically, Wisconsin is a very competitive state. Statewide elections, in particular, are often hotly contested with races often decided by factors unique to a single election rather than straight partisan alliances. Recognizing this situation is crucial when analyzing statewide elections as a basis for determining redistricting boundaries. Any single election results brings its own set of circumstances that may not be applicable in predicting the likely outcome of legislative races. To overcome the bias of any one race, a variety of different types of elections were categorized and combined to develop a broader index of voting patterns for democratic candidates.

ADC1 clearly depicts the traditional voting patterns across the state of Wisconsin. Readily apparent are areas of high democratic vote percentages (shown in green tints) of the state's major urban areas, including Milwaukee and Racine-Kenosha in the southeast; Madison in the south-central; Stevens Point and Wausau in the central, and the larger rural area in the west-central and northwest areas of the state.

Likewise, areas of strong Republican vote tallies (shown in red tints) appear in the suburban southeast; the east-central Fox Valley area, and the rural areas through the center of the state extending from the Illinois state line north to the Michigan state border.

Each cell or area on ADC1 represents a civil town, village or city. The very regular square "cells" covering most of rural Wisconsin are a reflection of the 6-mile by 6-mile square townships laid out in the state's original public land survey system. In many instances, civil towns cover the identical area as the surveyed township.

The full range of partisanship in Wisconsin can be seen in the Milwaukee County inset. From the dark green of the extremely Democratic inner-city Milwaukee, to the dark red of the rock-ribbed Republican Milwaukee suburbs, a distinct range of solid and consistent political affiliations is clearly evident.

Data Selection

The data for this map are based on 11 statewide elections held from 1982 through 1990. Election races tallied include the following: governor 1982, president 1984, U.S. Senate 1986, governor 1986, attorney general 1986, state treasurer 1986, president 1988. U.S. Senate 1988, governor 1990, attorney general 1990, and state treasurer 1990. To create the mapped index, the total Democratic vote cast in the 11 races was divided by the total vote (both Democratic and Republican) for the 11 races. This particular methodology was chosen over the alternative of averaging the 11 individual percent Democratic tallies, so races with a higher turnout would add more significance to the overall averaged index.

In developing the index for partisan performance, both parties had the opportunity to request that certain races be added to the process, and the races for attorney general and state treasurer were added at the request of the Republicans. During later federal court hearings on the redistricting plan, Republican lawyers attempted to refute the Assembly Caucus Index by pointing out that it did not include several races for the secretary of state. The argument was overturned because both races involved a popular incumbent who was easily reelected without serious opposition, and because the Republicans had ample opportunity to request the addition of these races if they thought they were relevant.

Conclusion

ADC1 was an important tool in developing the 1992 legislative redis-
tricting plan, and in the successful defense of that plan in federal court. Since redistricting, ADC1 has been adopted by many political organizations as a standard measure of partisan-ship, and is the measure of democratic performance most widely quoted by the Wisconsin news media in election coverage.

Acknowledgement

Assistance with adapting the design and production of ADC1 and the Milwaukee County inset for inclusion in the URISA Journal was generously provided by Jerome Gill Sullivan, Land Information and Computer Graphics Facility, University of Wisconsin-Madison.
In this issue . . .

This issue presents three book reviews related to spatial data and information technology.

In keeping with the URISA '94 Conference theme, "Integrating Information & Technology: It makes sense," two of the book reviews focus on issues related to financial aspects of spatial data, information technology and GIS. While current data and technology hold great promise, we are still on the verge of turning that promise into financial gain or even savings.

*Profiting From a Geographic Information System* is one of the first books to focus on the use of GIS in the private sector. While the title calls attention to the profit aspect of GIS, the book goes far beyond this issue. Rebecca Somers' review highlights some of the key discussions addressed in the book.

John Behrens' review of *Urban Finance Under Siege* focuses on the financial plight of urban governments and the relationship of the financial issues and land data and information technology. This review is a departure from the technology-oriented reviews in that it first examines the problem and then discusses how technology and data may help solve it.

The third book addresses a topic of great discussion and debate—the National Spatial Data Infrastructure (NSDI). Nancy von Meyer reviews the book produced by the National Research Council Mapping Science Committee, *Toward a Coordinated Spatial Data Infrastructure for the Nation*.

Editorial Intent

The Reviews section of this journal provides critical reviews and information concerning publications, information sources, and software in the field of urban and regional information systems.

General review categories include: Books, Publications, Videos and Software. Software reviews are of three types: "In-Depth," "Head-to-Head," or "From the Inside."

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.

For complete submission guidelines, see p. 105-106.

Rebecca Somers
Peter Van Denmark
Profiting from writing about "profiting from a GIS," the foremost accomplishment of this book may be that of recognizing a market opportunity and being the first to address it. Business's interest in GIS is very high at the moment, yet applicable information is scarce. While GIS has become prevalent in the public sector, its use in business is lagging. This book and its individual authors recognize this situation and offer various explanations for it. More importantly, however, they attempt to describe GIS in business terms and applications. The establishment of this basic understanding is the first step in accomplishing the required technology transfer.

Overview
The book's editor states that it is intended to be read by everyone, that it will give you a competitive advantage, and that it will teach you how to "drive" a GIS without explaining what is happening 'under the hood.' He also cites forecasts that "GIS will be a $4 billion industry by 1996, with 17 percent of the growth occurring in the private sector." Later, he predicts that those companies who first take advantage of GIS technology will become "highly successful, famous and rich." That sounds good to me.

The book has five parts. Part 1 is a Technology Overview, introducing GIS concepts and capabilities. Part 2 presents Vertical Industry Applications, and consists of discussions of the use of GIS in selected business sectors and markets. Part 3 is entitled Data Sources, but also deals with a number of concepts and issues. Part 4 addresses the Future of GIS, presenting a variety of perspectives. The final part of the book presents resources for further information.

Technology Overview
The first section, "Technology Overview," consists of two chapters. GIS Fundamentals include discussions of what a GIS is, what it does, and how it works; the structures and functions of GIS, including data, hardware and software functions; and organizational and institutional factors. (This last topic seems to stray into the territory of the next chapter.) The discussion is very generalized and does not really present the concepts in a business context. Furthermore, it attempts to talk its way through all these graphic, geographic and geometric concepts without the aid of sufficient graphics.

The second chapter, GIS Capabilities, Uses and Organizational Factors, backs up a bit in a discussion of GIS functionality that was started in the first chapter. It would have been better continued there. At least this chapter uses graphics to illustrate the concepts. A good discussion of the organizational setting and the realities of implementing GIS technology and balancing current needs with corporate IS/IT enterprise directions follows.

While GIS concepts and capabilities is a critical place to start, the text in this section is somewhat disappointing. Later authors provide better explanations of some of these concepts.

Vertical Industries Applications
The "Vertical Industry Applications" section presents snapshots of GIS usage in selected industries. These include financial institutions, insurance, real estate, manufacturing and packaged goods, retail trade, health care, transportation and navigation, mobile telecommunication, and utilities. Each chapter discusses the uses of GIS in that industry, and some also discuss important issues in the implementation of GIS technology for those types of applications. Generally, the applications in this section are pretty terse, making use of limited GIS capabilities. This illustrates the editor's and authors' comments that business lags behind government in effectively using GIS. On the other hand, the best way to introduce people to GIS is to start with the simple, and most obvious, applications. Some of the maps in the various chapters assist in this regard, but many chapters are lack-
ing adequate graphics. In addition to giving the reader an industry-by-industry perspective on GIS, there are some good insights and graphics in various chapters.

The first chapter in this section, **Financial Institutions**, presents some excellent insights and examples that apply to many industries. Simple maps with appropriate geographic orientation and information are presented to illustrate various applications. (This would not be so remarkable were it not for the absence of good maps in other applications chapters.) The author, Lew-Jean King, also employs some very effective techniques for conveying concepts concerning GIS use and data that would have been very helpful if extended to other applications and industries. He develops a matrix illustrating bank department functions with respect to the GIS applications they would perform, the required GIS analytical functions and the required data. He also presents a spatial decision support model that incorporates GIS.

Likewise, in the chapter **Real Estate**, Gil Castle develops a three-dimensional matrix for matching potential GIS application with products, services and industry players in the real estate industry. A similar model would be useful for explaining other industries. He also makes some very insightful observations in this chapter that apply to most industries. He discusses the importance of integrating GIS into applications and executive information systems, and the fact that GIS will not attain its potential until two-way communications with other software tools and databases can be established.

Hal Reid’s chapter, **Retail Trade**, presents GIS applications concepts and issues that also apply to most other industries. His introduction to GIS is in business terms, accentuating the strategic and tactical concerns of any business. The maps he presents for illustration are excellent. They illustrate the concepts, data and application functions and are easily comprehensible. As with the **Financial Institutions** chapter, much of his discussion could be applied to most industries and applications, and is clearer and more to the point than some of the intentionally generic material at the beginning of the book.

Ann Badillo’s chapter, **Transportation and Navigation**, discusses some cutting edge applications, that are attempting to leverage the latest technology to improve transportation systems, operations, and conditions. She discusses the integration of communications, navigation and information systems technologies in two intelligent vehicle highway systems: Project Pathfinder and TravTek.

Joyce Rector’s chapter, **Utilities**, ends this section with a very informational account of the use of GIS in one of the industries in which it has been used the longest—often known as AM/FM. From this perspective, she conveys some important insights and wisdom from which GIS practitioners in other industries could learn much. This chapter is the only place in the book that cost/benefit and financial matters are addressed in any depth. Her discussion of the cost/benefit calculation process and the implementation decisions that are based on its outcomes could serve as a useful guide for many businesses. This paper is also well referenced.

### Data Sources

The section on “Data Sources” goes far beyond just discussing sources. It includes very cogent discussions of the role of data in a GIS: data characteristics; how to select, design and obtain databases; and critical issues in GIS implementation. Much of this text is better than in **GIS Fundamentals**, and might more properly replace that chapter—putting the better ideas “up-front.”

Don Cooke, author of the chapter, **Spatial Data for Business**, is one of the foremost GIS data pioneers. He offers a very insightful and straightforward discussion of spatial data characteristics and structures; the historical development and use of public geographic data; and data characteristics and related issues such as scale, resolution, formats, currency and delivery media. He also addresses the important issue of geocoding. In the rush to implement GIS, his aspect of database implementation and use is often overlooked. Government users have learned about geocoding the hard way. Most potential business users do not even know that there are critical, potentially large issues involved in “loading address data.” Hopefully they will learn from Don Cooke’s discussion. This chapter also presents data sources and other information on GIS and data.

Dean Snoecker contributed the chapter, **Attribute Data for GIS in Business**. He provides some of the book’s best discussion of the business perspective on data and its use. He introduces the concept of data vs. information vs. knowledge and puts this classification of databases in context with respect to combining a business’s proprietary information with outside information. He discusses the concepts and issues involved in selecting data, including data characteristics, data quality, packaging, services, expertise, and dealing with irregular geography.

Nora Sherwood Bryan wrote the chapter, **Government and the Business of Data**. As with the other authors of this section, she provides some excellent discussion regarding GIS data concepts. She discusses geography, scale, geographic units and the types of applications suited to these different data types. This is the first place in the book that the concepts of differing geographic scales and
their related geographic units, descriptions, structures, and application are actually illustrated. While discussing data sources and data acquisition, she also provides a reminder of the issues involved in government sales of data—matters that are far from clear at the moment. This chapter also furnishes a good directory of data sources.

Warren Ferguson provided the chapter on Conversion, meaning data conversion. He provides an excellent presentation of different approaches to GIS database creation and their implications. He describes the data conversion process and relates it to the entire GIS implementation process, including identification of data and GIS requirements and decisions regarding data scales, formats, and sources. He makes the important point that GIS have traditionally been developed through the long and expensive approach of creating large amounts of high resolution data, but that this approach is probably not suitable for many business applications (nor for many government applications in these days of fiscal restraints). Alternative data conversion and GIS implementation approaches are presented. He also discusses geographic data concepts and issues such as accuracy, connectivity and structure, database design and data translation.

The Future of GIS

Part 4 presents various perspectives on “The Future of GIS,” divided into one chapter of essays by representatives of the private sector of the GIS industry, and one of essays by those in the academic sector. Aside from the discussions in the “Data Sources” section, these short papers are some of the most interesting in the book. Peter Van Demark discusses the future of desktop mapping and the migration of GIS power to the user’s desktop. Ginger Juhl outlines trends in the industry, demonstrating how evolution in the computer industry is sending waves of change into the GIS industry and the widespread use of GIS. She addresses the important issue of GIS integration into larger IT infrastructures and discusses the effects of GIS on the way businesses will operate. John Antenucci addresses the issues of change. He also discusses the issue of GIS integration, and the fact that GIS will rightly become a data-handling engine within other applications. Peter Croswell addresses technology in the context of implementation and changing organization and business processes. He supports the points made by other authors that GIS should not be considered a niche, but a part of the larger technology and data infrastructure in an organization. Tony Buxton and John O’Callaghan take this concept of GIS integration even further by presenting a scenario in which a business person accesses the power of GIS through function buttons and information bubbles within the context of his application or decision-making process. The chapter also contains papers on knowledge-based systems, the growth of GIS use, data modeling and access, mass-market education, and enterprise approaches.

This chapter is right on the money—but bears little resemblance to some of the simplistic applications discussed in the vertical industry section. This gap makes the problem clear: the “future” is here now. The technologies and uses that are described in the GIS futures section (defined as five-year predictions) are critical elements of any GIS planning, acquisition, implementation or use decisions being made today. Business use is definitely lagging.

The second chapter in this section presents the academic perspective with seven essays from individuals in three countries. Michael Shiffer and Lyna Wiggins provide a paper, The Union of GIS and Multimedia, that highlights current technology and research that are paving the way for integrating data sources and forms, and related technologies with GIS. They discuss the potential for direct manipulation interfaces, multiple representations of information in different contexts, GIS modules that “snap together” like Lego®, and inter-applications communication. They also discuss the prospects of new technology such as portable GPS, PDA (personal digital assistant), wireless communication, and computer supported cooperative work (CSCW). Through this short paper, it is possible to envision the big changes that will take place in the way we interface with technology and incorporate it into our working environments. David Grimshaw echoes the prediction of many authors that GIS will become completely absorbed in many applications environments and larger corporate-wide information systems infrastructure.

The other papers in this chapter are sometimes out of step with the general trends of GIS development and use cited by other authors. The academic papers regarding GIS implementation and organizational issues appear to be based on the traditional (and soon to be out-moded) concept of large, stand-alone GIS. Perhaps the academic community does see the future of GIS differently from the business community, but these academic papers are a small, and not particularly representative, set. The NCGIA Centers—USC at Santa Barbara, SUNY at Buffalo and the University of Maine at Orono—were not represented, nor were other active GIS centers such as Ohio State University, University of Wisconsin–Madison, and the University of South Carolina, to name but a few.

GIS Profits and Prophets

Even though the title of the book refers to profits, there is little information about specific cost, benefits,
savings, cost reductions, income and profits accruing to the use of GIS. Most authors offer the same statements offered throughout the industry regarding improved operations, decision support, enhanced planning, reduced costs and saving. Many authors cite the price of various hardware arrangements. Some calculate savings in a per-unit time or dollar amount. Total and long-term savings and financial benefits are sometimes derived from these. The problem, of course, is that there are not enough historical data on GIS in business—that are public information at least. (If a company were making significant amounts of money due to their use of GIS, they may not be inclined to share their secrets with their competitors.) Even in the public sector, where GIS has been implemented for the past 20 years, actual figures concerning cost/benefit and savings are hard to come by. Still, the dollar factor is at the bottom line of every business person's decision to pursue GIS.

Although many authors cite data and system interfaces as future considerations, the book does not really address the role of GIS in systems integration. And that is surely where much of its future lies. Many leaders in the GIS field predict that GIS will disappear as stand-alone systems and become modules of applications systems and larger IS/IT and data management infrastructures. The omission of this point as a basic concept in the book would give the reader the impression that it is merely a side issue, or a matter of adding more in-depth discussion. In reality, systems integration, at a variety of levels, is a cornerstone of GIS for the future. While in-depth discussion would not have been appropriate in this book, the matter should have been pulled forward as a key concept. A perspective from a systems integrator would have been appropriate in the future trends section.

While the information contained in the book is very good, the editing leaves a bit to be desired. Most chapters appear to be written as isolated papers, with only editorial notes inserted for reference to other chapters of the book. Thus, much material is presented over and over again, such as the definition of GIS, data characteristics, the historical development and use of GIS, growth of the GIS industry, hardware component costs, basic GIS functional definitions, etc.—matters that should have been handled in Part 1. The lack of common terminology, concepts and a framework within which to present the applications viewpoints may lead a reader to perceive contradiction among authors. There is some, but the most apparent contradiction stems merely from the author's different ways of explaining things. While it's true that, taken in total, the book gives the reader the advantage of many viewpoints, the text could probably be reduced by at least a quarter with more careful editing.

A good reason for the repetition may be that it is expected that the reader will read selectively, and thus will be sure to acquire the necessary basic information. This might be possible, were it not for the fact that many of the basic concepts and applications functions are scattered throughout the book, and the casual reader would miss some important matters.

Many of the most insightful comments and cogent explanations are buried in the vertical applications and the data sections. In particular, some of the best discussions of GIS systems and data integration, GIS implementation, cost/benefit analysis, and the business person's perspective were voiced by authors of specific chapters not particularly designated to address those subjects. For example, how would a casual reader know that the best expression of a business person's view of GIS is contained in Hal Reid's chapter on Retail Trade? Or that the only real discussion of cost/benefit determination was in Joyce Rector's chapter on Utilities? If the reader were in the insurance industry and read only the "Technology Overview" section, the chapter on insurance, and perused the data source listings, he or she would have missed a lot of information that was very applicable.

Regarding more minor matters of structure and organization, there is not sufficient standardization of format, such as heading formats, enumeration (point, bullets, numbers), paragraph indicators, pull-quotes, minimum map presentation standards (some maps are just clusters or polygons with no identities—geographic or otherwise). While it is desirable to preserve the style of each author's paper, some standardization of structure and format would assist the reader throughout the various chapters of the book.

The book also needs more graphics—or at least in the right places. Most of the authors use graphics and maps extensively in their writing engagements, and know that the best way to present introductory information about GIS is through example. Yet somehow, the graphics didn't make it to the book. While the book laments the slowness of the introduction of GIS into business, it contributes to this situation by not presenting the most needed ingredient—good maps and examples.

Excellent information resources are provided at the end of the book, as well as within a few chapters. Given this book as a starting point, these references point the way to more in-depth discussion of various aspects of GIS, business applications and data.

Summary
This book goes on the "must-read" list, minor faults notwithstanding. It
provides a good (and maybe the only) overview of the applications and issues related to the use of GIS in a business environment. Great nuggets of information and wisdom are found in many of the chapters, and these insights often apply to all industries and aspects of GIS, not just the one the author is discussing—but you need to read the whole book to find them. On the other hand, the differences in authors' viewpoints, terminology, assessment of the GIS industry, predictions, and ideas concerning the state-of-the-art in business applications of GIS pretty well reflect the current state of affairs. On the balance, reading (or skimming) through the book will give you a good idea of the diversity of "the" GIS community and the myriad opportunities for using GIS in business.
“The Green Book” report of the National Research Council Mapping Science Committee was released in the spring of 1993, around the time of the Geo Data Policy Forum. Just as the National Research Council Reports in the early 1980s set the tone for debate about land information systems in the 80s, this publication may be the first of a series that set the debate for spatial data policy in the 1990s.

The question the committee addresses is what could be done better or more efficiently if the content, accuracy, organization and control of spatial data were different? It describes where we have come from, where we are, and some of the impediments in current systems that keep agencies from moving toward new solutions. Given this evidence, the committee concludes that the nation needs an improved National Spatial Data Infrastructure (NSDI) and presents a federal perspective on the principles and components of an improved NSDI.

The Mapping Science Committee has defined a National Spatial Data Infrastructure as a process for assembling and disseminating geographic information.

The National Spatial Data Infrastructure is the means to assemble geographic information that describes the arrangement and attributes of features and phenomena on the Earth. The infrastructure includes the materials, technology, and people necessary to acquire, process, store, and distribute such information to meet a wide variety of needs.

Using the analogy of the national transportation network, this means that the National Spatial Data Infrastructure corresponds to the gas taxes, construction methods, contracting procedures, and management methods for highways. The NSDI is not the traffic, pavement, and rights of way.

The committee examined the current state of affairs of spatial data in many federal agencies. These results are presented in terms of current data collection efforts and impediments to the NSDI. The impediments are largely a lack of leadership and a lack of coordinated direction. One conclusion from this section of the report is that government agencies need to change the way they do business and accept the true revolution that comes from using geographic data as an integrating concept in databases, organizational structure, and processing. The NSDI is not about putting a GIS terminal on every desk, it is about changing the way incentives and operations are managed and implemented in government. In failing to recognize that the NSDI is not “just another computer system,” agencies are erecting additional impediments.

Candidates for the NSDI are identified in the report. Specifically the landownership fabric, street centerlines and address, and wetlands are examined as potential NSDI data sets. The reasons for including these data in the NSDI, and examples of how these data should be changing the way business is done, are presented. Of the three examples, the landownership fabric is the weakest presentation and the wetlands is the strongest. The wetlands data study addresses the committee’s question of what could be done better and illustrates why government needs to change to take full advantage of spatial data. From this discussion, the report makes it clear that the solution to the NSDI is not “just do it.” There are clear issues of semantics, mandates, and incentives that must be resolved before the nation and government agencies can respond effectively.

One of the key components to an NSDI is data sharing. The Committee presents a new vision of data sharing based on automating data at the point it is produced. Partnerships, in the true sense of partnerships, are proposed for local governments, state agencies, and the private sector to contribute to the NSDI. Data donors are encouraged to participate through economic support and data credits. This is different from current data production

Nancy von Meyer is vice president at Fairview Industries in Middleton, Wisconsin.
arrangements where interested parties pay a federal agency to collect and compile the data they need. This is a partnership where the most appropriate participant collects and compiles data to agreed-upon standards, and other participants provide support to the collector. To demonstrate the difference, the committee might have called the new arrangements data relationships rather than partnerships, reflecting the longer-term, more-cooperative nature of these arrangements.

The committee has identified four principles for an improved NSDI. These are:

- **Availability**—The results of the NSDI processes must serve everyone, not just federal agencies.
- **Ease of Use**—The NSDI must be easily applied by all government agencies, private sector organizations, and citizens.
- **Flexibility**—The NSDI must be able to respond to emerging technologies.
- **Foundation for Other Activities**—The NSDI is a springboard for fostering sustainable development and growth.

The Report concludes that national goals for the NSDI should be established at the federal level to assure integration. The Federal Geographic Data Committee (FGDC) should continue as a coordinating body with expanded emphasis on completing standards (content, quality, performance and exchange) and on developing a series of incentives for data sharing. The federal level also should foster ready access to information and an enhanced cross-cutting data-sharing program.

One of the real gems in this report is Appendix A, a study on wetlands. The policy implications of defining and characterizing wetlands is presented from local, state and federal perspectives. The information requirements to meet that policy are described as well as impediments to information development. Based on the results of the wetland case study, an information diffusion model is presented that relates policy to information. The NSDI community would be well served to have this type of careful study for all components of the NSDI.

A lingering question the report does not address completely is: where are the benefits? Returning to the transportation system analogy, the primary benefit of the national transportation systems was national defense. What are the specific advantages of spatial data moving openly and freely among all agencies? The Mapping Sciences Committee discusses some benefits related to competitive advantage, economic leverage, and general good government. Are these tangible enough benefits to compete for funding with hungry children, health care and unemployment?

As an example, the Wisconsin Land Records Program grew from a benefit basis. The Larsen Report (1978) established that an average citizen in Wisconsin paid approximately $17 per year for land records in 1978 dollars. This number became both a rallying and selling point for land professionals in the state. By using spatial data as an integration concept, and leveraging technology into local government, land records management became more efficient.

The Mapping Sciences Committee report sets the stage for the spatial information professional community to address the issues of semantics, mandates, impediments and incentives for a national spatial data infrastructure. The atmosphere for reinventing government is ripe for a mature NSDI.

**Reference**

Generalizations have always been risky, and this book contains its share, some even bolstered by indicative figures. Moreover, a primary fiscal source in the American urban landscape—the property taxes—adds its own challenge to anyone sturdy enough to assay its present significance. Property taxes are inherently local, and thus diverse, in impact, administration, and effectiveness, whether they are local levies as in the United States (whereover 60,000 local governments can and do impose them, extracting three of every four local tax dollars nationwide), or state taxes (42 of these United States have the levying power, which some actually use, but only for something near 1 percent of total revenue annually). Indeed, property taxes wickedly attract sorely tried savants: almost anything you say about them will apply somewhere, and almost nothing you say about them will apply everywhere. They are local and plural. Speaking of the property tax, as though we have finally found the generic model, can obscure but never really hide the variety that property taxes exhibit. The one in Waukegan is different from the one in Dallas, and both are different from the one in Manchester, New Hampshire.

In important respects, cities (and not only American cities) also differ a lot among themselves, similarities notwithstanding. In this book, then, many ingredients condition and complicate analysis of the U.S. urban environment, even when it’s under siege, and especially at a time of profound technological change.

The Authors and the Setting
Seven intrepid economists nevertheless contributed a chapter each to the book, two of them also serving as co-editors. These latter are Thomas R. Swartz and Frank J. Bonello, each of whom has advised various governments, and both of whom now teach at Notre Dame.

The authors, all of whom have veered back and forth between the “real world” of government policy jobs or consulting assignments, and that other “real world” of academe, include Dick Netzer, the long-time, yet very-contemporary seer from New York University; Katharine C. Lyall, president of the University of Wisconsin system, economics professor at that university, and a HUD official in the Carter administration; James R. Follain, economics professor and chairman of the department at Syracuse University; Andrew Reschovsky, professor of agricultural economics and public affairs at the University of Wisconsin; and William K. Tabb, professor of economics at Queens College, City University of New York.

Swartz opens by advocating the need to “rethink” urban finance, acknowledging that the Tiebout effect (fleeing local governments that do not please you) does not always apply. The unemployed and many in minority groups are trapped in today’s cities because an unenlightened society inhibits their mobility.

Swartz also notes that the years when federal grants supplied 10 percent (in the 1950s) to 25 percent (in the 1970s) of local spending are now over, with state aid now an inadequate source of assistance to local governments. He then rues the decline in new household formation during the 1990s, following the boom years after World War II and the “baby bust” decades of the 1960s and 1970s. He places such happenings in the context of research by Gregory Mankiw and David Weil, who estimated, in Regional Science and Urban Economics for May 1989, that the real price of housing might decline by 47 percent between then and 2010. This, for some, suggested related disaster for the residential base of property taxes.

For commercial realty, the Economic Recovery Tax Act of 1981 seemed a genuine boon, first by cutting personal income taxes for upper echelons of income, leaving a wake of after-tax dollars to invest, then by easing tax obligations applicable to investments in commercial property. Not surprisingly, total mortgages for commercial realty rose from $225.5 billion in 1980 to $555 billion in 1986. The rude awakening came with the Tax Reform Act of the same year.

John O. Behrens is an attorney and former tax assessor whose background includes various assignments in public finance in the United States and other countries, including Liberia, Sierra Leone, and the Philippines. He is past president of the Institute for Land Information. Currently he is a visiting fellow at the U.S. Advisory Commission on Intergovernmental Relations (ACIR).
The TRA ended the special tax advantages for investment in property development, and also increased the marginal tax rates involved by classifying rental income in the "passive income" category. TRA deemed taxable. Such influences on residential and commercial property may well have had depressing effect on property tax bases. Follain, for one, however, cited 1991 articles disputing Mankiw and Weil, then gave his own "best guess" that the value of the owner-occupied housing stock would increase by more than 25 percent in real terms during the 1990s.

In any event, the decline in federal spending for local programs, cited in various ways by Swartz, Lyall, and Tabb, was certainly enough to stimulate state and local thinking about the respective potentials among available revenue sources. Tabb noted that federal grants dropped from 27 percent of state and local spending in 1978 to 17 percent ten years later. Lyall pointed to a potent array of percentage losses among federal grants to cities between 1980 and 1991: general revenue sharing down by 96 percent, local public works down by 57 percent, city water and sewer grants down by 47 percent, energy conservation down by 45 percent, and mass transit down by 48 percent.

Meanwhile, Lyall noted, Baltimore, Chicago, Detroit, and Philadelphia each lost 35 percent of their manufacturing employment bases during the 1970s. Cities frequently said goodbye to jobs.

Facts, Folks and Technology

All in all, there's enough bad news in what the economists report to fuel genuine concern. But the saving grace, coming through as much by osmosis as by analysis, is the absence of a gloom consensus, except for Lyall and Tabb. Lyall expects urban decay, unless dramatic shifts occur among levels of government to diversify workforces, renew infrastructures, and minimize unfunded mandates. Tabb sees the imperative need for a more globally oriented, but still inclusive, realistic kind of liberalization to effect justice and opportunity for groups either cynical or without hope, or both. Reschovsky thinks people will continue to resist local tax rate increases, believing (correctly or not) that too much of the increased revenue goes for spending on marginal groups.

Dick Netzer, more phlegmatic, dutifully notes the historic rise and subsequent decline in property taxes, which are now again on the rise. They did indeed account for nine out of ten local tax dollars up to World War II and still bring in three of every four now. In fiscal 1992, the annual local yield reached $172 billion, or 26 percent of all local revenue, down from 43 percent in 1957, but respectable nonetheless. Among all taxes of all types levied by all governments in the United States, local and state property taxes accounted for 14.6 percent in 1992, an increase from 12.3 percent 10 years earlier. A collateral happening, deemed beneficial by many, is the more frequent occurrence of more diversified local revenue structures that include any nonproperty taxes (such as local earnings taxes) and nontax revenue (such as charges for current services), that may be at once legally available and economically viable.

Netzer and Reschovsky point as well to one important influence currently a significant factor in assessment administration and property taxation, namely, technology. They then mention a curious deficiency (discussed later), not in property taxation itself, but in the unpleasant job of providing the information necessary to administer property taxes efficiently. Andrew Reschovsky had also called attention to state court litigation, where decisions mandated timely assessments, at the value level specified in state law. During the '80s those "timely" assessments lingered, even after market values declined. This gave the jurisdictions involved relatively high valuations while they re-examined revenue policy.

The important influence Netzer and Reschovsky both identify is technology, a development enhancing numerous aspects of urban life, but particularly crucial with regard to property taxes. Their history is replete with instances of barely adequate, even less than adequate, administration. In manual (pre-computer) environments, property tax assessors, even when competent, had impossible jobs. Assessors were to produce a timely, annually assessed value for each taxable property, which was to be uniform with the assessed value for each of the thousands of other taxable properties. Frequently the assessors failed, and expeditors such as fractional assessments and multi-year assessment cycles became common. Given his choice, Dick Netzer would probably get rid of all expeditors and tax land more than improvements, in a coldly dispassionate endorsement of Henry George's thesis (thesis, that is, not panacea). After noting that relatively low land taxes have no effect other than to make present owners richer, Netzer bluntly states that, for cities at least, "the best of all possible revenue systems" would include only "taxes on the value of land to finance public goods, and marginal cost-based user charges for everything else." The homely problem, however, is that existing property taxes too often undertax land value, and in any case do not easily fit as user charges.

That millennium, alas, remains distant, but assessors today in many jurisdictions benefit from modern technology and actually do a better job. Computer-assisted mass appraisal (CAMA), whereby the assessor, with computer, estimates each property value (dependent variable) on the basis of attributes (indepen-
dent variables) of sold properties, has become a fixture in thousands among the 13,600 local assessing offices in the United States. With CAMA, annual, appraisal-type reassessment of each property becomes a realistic objective. And in some jurisdictions it's already an actual achievement.

A second "high-tech" infusion within assessing offices is the geographic information system (GIS), the modern answer to an assessor's need for a system of accurate cadastral maps. In the past, that need was more frequently ignored than recognized. All indications are that the future will be different. Folliol is probably the most optimistic about that future. He thinks that income and population both will continue to grow in the 1990s. He also thinks pressure will increase to limit the extent to which high-income homeowners can deduct mortgage interest for federal income tax purposes. He agrees with that targeted limit, pointing out that in 1989, when this tax subsidy exceeded $80 billion, 30 percent went to those with adjusted gross incomes higher than $75,000 and 55 percent went to those with AGI's over $55,000.

Reschovsky places some hope in demographics, expecting fewer young workers in cities of tomorrow, along with consequent labor shortages and higher wage rates. He also believes it possible that health care may become a strictly federal responsibility, and that lower market values for housing may in the long run reinvigorate local economies.

All of the foregoing places in stark relief the "curious deficiency" referred to earlier and noted by Netzer and Reschovsky alike, namely, that there exists "no comprehensive national data on trends in the market value of property. The most recent data for a broad sample of the U.S. cities date from 1981." So says Andrew Reschovsky. He is speaking of Taxable Property Values and Assessment-Sales Price Ratios, Volume 2 of the 1982 Census of Governments, still, 12 years later, the most recent source of internally consistent local, state, and nationwide data on assessed values, de facto, levels of assessment, coefficients of dispersion, and effective tax rates. In every fifth year beginning in 1957, the U.S. Bureau of the Census, often innovative and always objective, has enumerated samples of individual real property sales and individual real property assessed values to produce objective findings on property assessment, market value, and effective property tax rates applicable to real property in many local jurisdictions around the country. It is highly regarded, the only survey of its kind, public or private, in the entire country, yet it struggles for existence. Its results make possible the estimation of market value of real property for jurisdictions throughout the country, and for the country as a whole. Because its data are internally consistent, its findings are uniquely valuable for estimating magnitudes of real property market value, and changes affecting such value from one time period to the next, in and sometimes within or between cities, counties, states, and the entire country. Thus policy-makers, assessors, other government officials at local, state, and federal levels, and economists like those who wrote this book look to this Census Bureau effort as a tool for estimating magnitudes and changes affecting real property value (including those related to housing), property tax burdens (as indicated by effective tax rates), assessment performance, and de facto levels of assessment. Since 1982, however, the funds necessary to continue it in its comprehensive, most useful form, have not been adequate. Indeed, the Census Bureau cancelled the assessment-sales price ratio study portion entirely for the 1987 Census of Governments, and there is now fear that only a truncated version will emerge for the 1992 census publication, now in progress. Let's hope, instead, that the Census Bureau finds the funds to produce this information treasure for an information society eager to use it.

Concluding Comment

It is instructive to note that, despite provocative inferences and careful conclusions about the urban fiscal environment perceived to be under siege today, the authors betray a yearning for the even more incisive judgments that the technology of the information age can give them once it transforms the landscape more definitively. There are the several references to the need for more data, together with acknowledgments that new departures confront us (as Lyall notes in alluding to the "edge cities" like Las Colinas, Texas, that Joel Garreau identified).

Throughout the book there is an awareness that the inclusive society, with a better life for all, will continue to elude us unless we enlist everything available in our efforts to overcome the problems afflicting urban finance, under siege or not. The comprehensive improvement deemed necessary will likely involve a practical composite of old and new fiscal measures, all used to further the policy fundamental that all people count.

In that setting, the innovative, cost-effective technology essential for the information society is the kind of natural infusion that will bring the inclusive society within reach. Computerization in the service of more responsive fiscal instruments and more useful geographic information can be the abiding catalyst for that society's emergence.

Note

Opinions expressed are the author's and not necessarily those of ACIR.
IN MY OPINION/EDITORIALS
Archer, Hugh. “Expanding Affordable Public Access To Information: Where Do We Go From Here?” Vol. 5.2: p. 5.

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.
Lee, Jay. “Dynamic Delivery of Transportation Services with GIS.” Vol. 3.2: p. 3.


**FEATURES**


Bertocci, Roberta; Francesco Brunori, Paolo Canutti, 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.


Author Index/URISA Journal 101


Norman, Margaret. URISA Crossword Puzzle. Vol. 2.2: p. 73.


Tsui, Mary. “Data Integrity...G.I.G.O. Revisited.” Vol. 5.1: p. 81.


FEATURE MAP


REVIEWS
Books


Monograph
AURISA. “Toward the Implementation of a National Strategy for

Publications
Crane, Edward. Review of American Public Works Association, 

Research

Software


Video
Subject Index for Refereed Articles

<table>
<thead>
<tr>
<th>Analysis and Planning</th>
<th>Volume</th>
<th>Lead Author</th>
<th>Benefit/cost methods</th>
<th>Volume</th>
<th>Lead Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling</td>
<td>2; 2: 35</td>
<td>Watterson</td>
<td></td>
<td>4; 2: 20</td>
<td>Poe</td>
</tr>
<tr>
<td>Economic analysis</td>
<td>3; 1: 6</td>
<td>Chenoweth</td>
<td></td>
<td>2; 1: 11</td>
<td>Epstein</td>
</tr>
<tr>
<td>Routing</td>
<td>3; 2: 50</td>
<td>Tomaselli</td>
<td></td>
<td>3; 1: 33</td>
<td>Steger</td>
</tr>
<tr>
<td>Small area forecasting</td>
<td>3; 2: 3</td>
<td>Lee</td>
<td></td>
<td>4; 1: 45</td>
<td>Dando</td>
</tr>
<tr>
<td></td>
<td>2; 2: 35</td>
<td>Watterson</td>
<td></td>
<td>2; 1: 38</td>
<td>Roitman</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2; 2: 2</td>
<td>Grady</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2; 2: 16</td>
<td>Langran</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 1: 24</td>
<td>Denkers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1; 1: 27</td>
<td>Nyerges</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1; 1: 17</td>
<td>Parent</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3; 1: 20</td>
<td>Craig</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3; 2: 25</td>
<td>Nolan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 1: 10</td>
<td>Frank</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 2: 59</td>
<td>Blinn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2; 1: 11</td>
<td>Epstein</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3; 1: 14</td>
<td>von Meyer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3; 1: 43</td>
<td>Croswell</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 1: 32</td>
<td>Oursud</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 2: 59</td>
<td>Blinn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2; 1: 38</td>
<td>Roitman</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2; 1: 2</td>
<td>Hintz</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3; 1: 14</td>
<td>von Meyer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 1: 45</td>
<td>Dando</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1; 1: 17</td>
<td>Parent</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2; 1: 38</td>
<td>Roitman</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 1: 45</td>
<td>Dando</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1; 1: 7</td>
<td>Craig</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 1: 32</td>
<td>Oursud</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1; 1: 39</td>
<td>Clapp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4; 2: 47</td>
<td>Hazleton</td>
</tr>
<tr>
<td>Applications</td>
<td>1; 1: 17</td>
<td>Parent</td>
<td>Public access/freedom</td>
<td>4; 2: 32</td>
<td>Rudnicki</td>
</tr>
<tr>
<td></td>
<td>2; 1: 26</td>
<td>Kjerne</td>
<td>of information</td>
<td>2; 1: 2</td>
<td>Hintz</td>
</tr>
<tr>
<td></td>
<td>2; 2: 2</td>
<td>Grady</td>
<td>Research issues</td>
<td>2; 2: 35</td>
<td>Watterso</td>
</tr>
<tr>
<td></td>
<td>2; 2: 16</td>
<td>Langran</td>
<td>Systems evaluation</td>
<td>3; 1: 14</td>
<td>von Meyer</td>
</tr>
<tr>
<td></td>
<td>3; 2: 12</td>
<td>Armstrong</td>
<td>Systems design</td>
<td>4; 2: 7</td>
<td>Egenhofer</td>
</tr>
<tr>
<td></td>
<td>3; 1: 14</td>
<td>von Meyer</td>
<td></td>
<td>2; 1: 2</td>
<td>Hintz</td>
</tr>
<tr>
<td></td>
<td>3; 1: 20</td>
<td>Craig</td>
<td></td>
<td>3; 1: 25</td>
<td>Nolan</td>
</tr>
<tr>
<td></td>
<td>4; 1: 24</td>
<td>Denkers</td>
<td></td>
<td>3; 2: 35</td>
<td>Buyong</td>
</tr>
<tr>
<td></td>
<td>4; 1: 2</td>
<td>Frank</td>
<td></td>
<td>4; 2: 20</td>
<td>Poe</td>
</tr>
<tr>
<td></td>
<td>3; 2: 35</td>
<td>Buyong</td>
<td></td>
<td>2; 1: 26</td>
<td>Kjerne</td>
</tr>
<tr>
<td></td>
<td>4; 1: 10</td>
<td>Egenhofer</td>
<td></td>
<td>3; 2: 25</td>
<td>Buyong</td>
</tr>
<tr>
<td></td>
<td>4; 1: 24</td>
<td>Denkers</td>
<td></td>
<td>3; 2: 14</td>
<td>von Meyer</td>
</tr>
<tr>
<td></td>
<td>4; 2: 3</td>
<td>Egenhofer</td>
<td></td>
<td>1; 1: 27</td>
<td>Nyerges</td>
</tr>
<tr>
<td></td>
<td>2; 1: 26</td>
<td>Kjerne</td>
<td></td>
<td>1; 1: 39</td>
<td>Clapp</td>
</tr>
<tr>
<td></td>
<td>3; 2: 14</td>
<td>von Meyer</td>
<td></td>
<td>4; 1: 26</td>
<td>Beroggi</td>
</tr>
<tr>
<td></td>
<td>3; 2: 55</td>
<td>Buyong</td>
<td></td>
<td>2; 1: 16</td>
<td>Langran</td>
</tr>
<tr>
<td></td>
<td>3; 1: 14</td>
<td>von Meyer</td>
<td></td>
<td>2; 2: 16</td>
<td>Langran</td>
</tr>
<tr>
<td></td>
<td>1; 1: 27</td>
<td>Nyerges</td>
<td></td>
<td>4; 2: 32</td>
<td>Rudnicki</td>
</tr>
<tr>
<td>Databases</td>
<td>3; 1: 17</td>
<td>Parent</td>
<td></td>
<td>4; 2: 47</td>
<td>Hazleton</td>
</tr>
<tr>
<td></td>
<td>2; 1: 26</td>
<td>Kjerne</td>
<td></td>
<td>2; 2: 26</td>
<td>Watterso</td>
</tr>
<tr>
<td></td>
<td>2; 2: 2</td>
<td>Grady</td>
<td></td>
<td>2; 2: 7</td>
<td>de Neufville</td>
</tr>
<tr>
<td></td>
<td>3; 2: 35</td>
<td>Buyong</td>
<td></td>
<td>3; 2: 3</td>
<td>Lee</td>
</tr>
<tr>
<td></td>
<td>2; 2: 2</td>
<td>Egenhofer</td>
<td></td>
<td>3; 2: 50</td>
<td>Tomaselli</td>
</tr>
<tr>
<td></td>
<td>4; 1: 24</td>
<td>Denkers</td>
<td></td>
<td>4; 1: 45</td>
<td>Dando</td>
</tr>
<tr>
<td></td>
<td>4; 1: 2</td>
<td>Frank</td>
<td></td>
<td>4; 2: 59</td>
<td>Blinn</td>
</tr>
<tr>
<td></td>
<td>3; 2: 14</td>
<td>von Meyer</td>
<td></td>
<td>4; 2: 3</td>
<td>Egenhofer</td>
</tr>
<tr>
<td></td>
<td>3; 2: 3</td>
<td>Buyong</td>
<td></td>
<td>2; 1: 2</td>
<td>Hintz</td>
</tr>
<tr>
<td></td>
<td>2; 2: 16</td>
<td>Langran</td>
<td></td>
<td>3; 2: 25</td>
<td>Nolan</td>
</tr>
<tr>
<td></td>
<td>2; 2: 7</td>
<td>Rudnicki</td>
<td></td>
<td>3; 2: 35</td>
<td>Buyong</td>
</tr>
<tr>
<td></td>
<td>4; 2: 2</td>
<td>Egenhofer</td>
<td></td>
<td>4; 2: 20</td>
<td>Poe</td>
</tr>
<tr>
<td></td>
<td>4; 2: 26</td>
<td>Kjerne</td>
<td></td>
<td>3; 2: 25</td>
<td>Buyong</td>
</tr>
<tr>
<td></td>
<td>3; 2: 55</td>
<td>von Meyer</td>
<td></td>
<td>3; 1: 14</td>
<td>Nyerges</td>
</tr>
<tr>
<td></td>
<td>4; 2: 3</td>
<td>Beroggi</td>
<td></td>
<td>1; 1: 39</td>
<td>Clapp</td>
</tr>
<tr>
<td></td>
<td>2; 1: 26</td>
<td>Kjerne</td>
<td></td>
<td>1; 1: 56</td>
<td>Beroggi</td>
</tr>
<tr>
<td></td>
<td>2; 1: 17</td>
<td>Parent</td>
<td></td>
<td>4; 2: 47</td>
<td>Hazleton</td>
</tr>
<tr>
<td></td>
<td>2; 2: 2</td>
<td>Grady</td>
<td></td>
<td>2; 2: 6</td>
<td>de Neufville</td>
</tr>
<tr>
<td></td>
<td>3; 2: 50</td>
<td>Tomaselli</td>
<td></td>
<td>3; 2: 3</td>
<td>Lee</td>
</tr>
<tr>
<td></td>
<td>4; 1: 45</td>
<td>Dando</td>
<td></td>
<td>4; 2: 59</td>
<td>Blinn</td>
</tr>
<tr>
<td></td>
<td>4; 2: 3</td>
<td>Egenhofer</td>
<td></td>
<td>2; 2: 2</td>
<td>Watterso</td>
</tr>
<tr>
<td></td>
<td>1; 1: 27</td>
<td>Nyerges</td>
<td></td>
<td>1; 1: 39</td>
<td>Clapp</td>
</tr>
<tr>
<td></td>
<td>1; 1: 39</td>
<td>Clapp</td>
<td></td>
<td>4; 2: 56</td>
<td>Beroggi</td>
</tr>
</tbody>
</table>

104 URISA Journal/Subject Index
Guidelines for Manuscript Submission

REFEREED ARTICLES

1. URISA Journal welcomes manuscripts and accompanying graphics and illustrations on all topics that are germane to information systems, their evaluation and implementation. Articles must be based upon sound scholarship and provide information that is relevant to information systems and associated disciplines.

2. Since the review process requires that voluntary evaluators spend a significant amount of time selecting papers for publication, the submission of a manuscript marks the author's intention to publish in URISA Journal. Therefore, the simultaneous submission of a manuscript to other journals is considered unacceptable. Also, manuscripts previously published, either in a literal or approximate form, ordinarily cannot be accepted. Consult the editors if in doubt.

3. Submit three (3) copies of the manuscript to:
Kenneth J. Dueker, Portland State University, Center for Urban Studies, P.O. Box 751, Portland, OR 97207.

4. Be certain to retain a fourth copy for your own files. All submissions must be typewritten and double spaced on 8” × 11” paper. Leave a margin of at least 1” on all sides. Double space all material, including lengthy quotations, the abstract, notes, and references.

5. The cover page should include the title of the paper and the author's name and address. Also include, on this page, a two- or three-sentence biographical sketch summarizing the author's education, professional positions, current affiliation, and research interests.

6. Submit a one-paragraph abstract of approximately 100 words. Include the title of your paper with your abstract, but do not place your name on this page. Similarly, do not indicate your name on the first page of the manuscript.

7. To ensure anonymity in the blind review process, authors should be careful that the text not refer directly to their previously published works. In other words, avoid sentences such as: “The author's (Doe 1075, 1977) previous work revealed that.” If the paper is accepted, sentences can later be reviewed to reflect the above.

8. Within the text, references should be cited by using the author's name and year of publication. When using direct quotations, also include the page number(s). For example: Many employers and corporations have chosen to pursue a hands-off policy (Taylor 1915). "City planning and unified architectural design," according to Tunnard and Reed (1953, p. 131), "were lost to these new communities."

9. Multiple references in the text should be listed chronologically rather than alphabetically (Zube 1973; Jackson 1978; Tuan 1980).

10. If necessary, explanatory notes may be used. These should be numbered consecutively and must be included on separate pages at the end of the text.

11. First-order subheadings should be capitalized (e.g., FIRST-ORDER) Text should follow on the line below the subhead.

12. Second-order subheadings should be placed to the left margin on a separate line. The first letters of major words should be capitalized. The text should follow on the line below the subhead.

13. Third-order subheadings should be placed to the left-margin and followed by a period. The first letters of major words should be capitalized. The text should begin on the same line.

14. Long quotations (five or more lines of typescript) must be indented five spaces and double spaced.

15. All tables should be typed on separate pages. Indicate where a table or figure should be placed in the text by including notations such as (Table 1) or (Figure 1). Wherever possible, place these notations at the end of a sentence. Try to avoid sentences such as "Table 1 shows that..." Instead, use constructions that place the table or figure number at the end of the sentence:

   "The data indicate that all species were susceptible (Table 1)."

16. Any acknowledgement must be included at the end of the text. The heading for this page should be: Acknowledgements.

17. Authors will receive two copies of the URISA Journal issue in which their article appears. Reprints of articles can be ordered, at cost, by the author.

FEATURES AND REVIEWS

1. Submit two (2) copies of the manuscript and accompanying photos, graphics or illustrations to the appropriate section editor listed below. Be certain to retain a copy for your own files.

   All submissions must be typewritten and double spaced on 8” × 11” paper. Leave a margin of at least 1” on all sides. Double space all material, including lengthy quotations, notes, and references.
2. The cover page should include the title of the paper and the author’s name and address. Also include, on this page, a two- or three-sentence biographical sketch summarizing the author’s professional position, current affiliation, and interests.

3. Follow refereed article guidelines listed above for proper presentation of subheadings, quotations, tables and acknowledgements.

4. Contact the following section editors for details regarding specific manuscript formats for categories such as software reviews:

   - Features editor:
     Warren Ferguson, 10100 Reunion, Suite 715, San Antonio TX 78216
     Rebecca Somers, Somers-St. Claire, 3157 Babashaw Ct., Fairfax, VA 22031.

5. Authors will receive two copies of the URISA Journal issue in which their article appears. Reprints of articles can be ordered, at cost, by the author.

---

**Guidelines for Illustrations**

Photocopies of proposed illustrations, photos, or other graphic materials should be submitted with the text at the time of submission (three (3) copies for refereed section, two (2) copies for other sections). Include, on a separate page, a brief title and explanation for each figure.

**Line Copy**

1. Line copy (e.g., sketches, graphs, computer graphics) should be submitted on high-quality paper or as original drawings on illustration board or drafting paper in black ink.

2. Graphic material should not normally exceed 8 X 11 inches (21.5 X 28 cm), in order to prevent damage during mailing.

3. Line copy will often be reduced to meet space limits. Therefore, avoid extremely small print or complicated graphics (such as fine cross-hatching or dot patterns), which would not "read" at a reduced size.

4. Lettering on all copy should be clean and open. Use capital letters whenever possible.

5. All line copy should be identified with: (1) author’s name, (2) the figure number, (3) where applicable, artist credits.

**Photographs**

1. Submit only original glossy black-and-white photographs. (Color is acceptable only for map section.) 5 X 7 inch or 8 X 10 inch sizes are acceptable.

2. Avoid fingerprints, heavy pen or pencil marks, or damage from paper clips on the photographs.

3. Identify all photographs with: (1) author’s name, (2) the figure number, (3) credits (e.g., Photograph courtesy of Arizona State Museum, Jack Smith, photographer).

---

**Guidelines for Feature Map Submission**

1. The focus of the map section is on the image itself. We will consider both color and black-and-white submissions. You may submit: a complete map, an enlarged portion of a map, or both. The following criteria will be taken into account: visual quality of the map; amount and quality of information transmitted; and reproducibility of the map within the journal context.

2. We will accept printed maps or maps available in hard copy, provided they are clean and that you indicate the portion you wish to be published. Alternatively, we require glossy 8 X 10 color or black and white prints. Slides are acceptable but must be accompanied by a print. All submissions should be of high quality. This is especially important for computer-generated images, which can be difficult to reproduce.

3. Two to six double-spaced pages of text should discuss the map and its construction.

4. Include a separate page with the title, and author’s name and address. Also include, on this page, a two- or three-sentence biographical sketch summarizing the author’s professional positions, current affiliation, and research interests.

5. Any acknowledgement must be included at the end of the text under the heading: Acknowledgements. If your submission is a copyrighted image, please contact the map section editor to make special arrangements.

6. Submit map materials and text to the section editor:

   Ted Koch
   Wisconsin State Cartographer
   160 Science Hall
   550 N. Park St.
   Madison, WI 53706