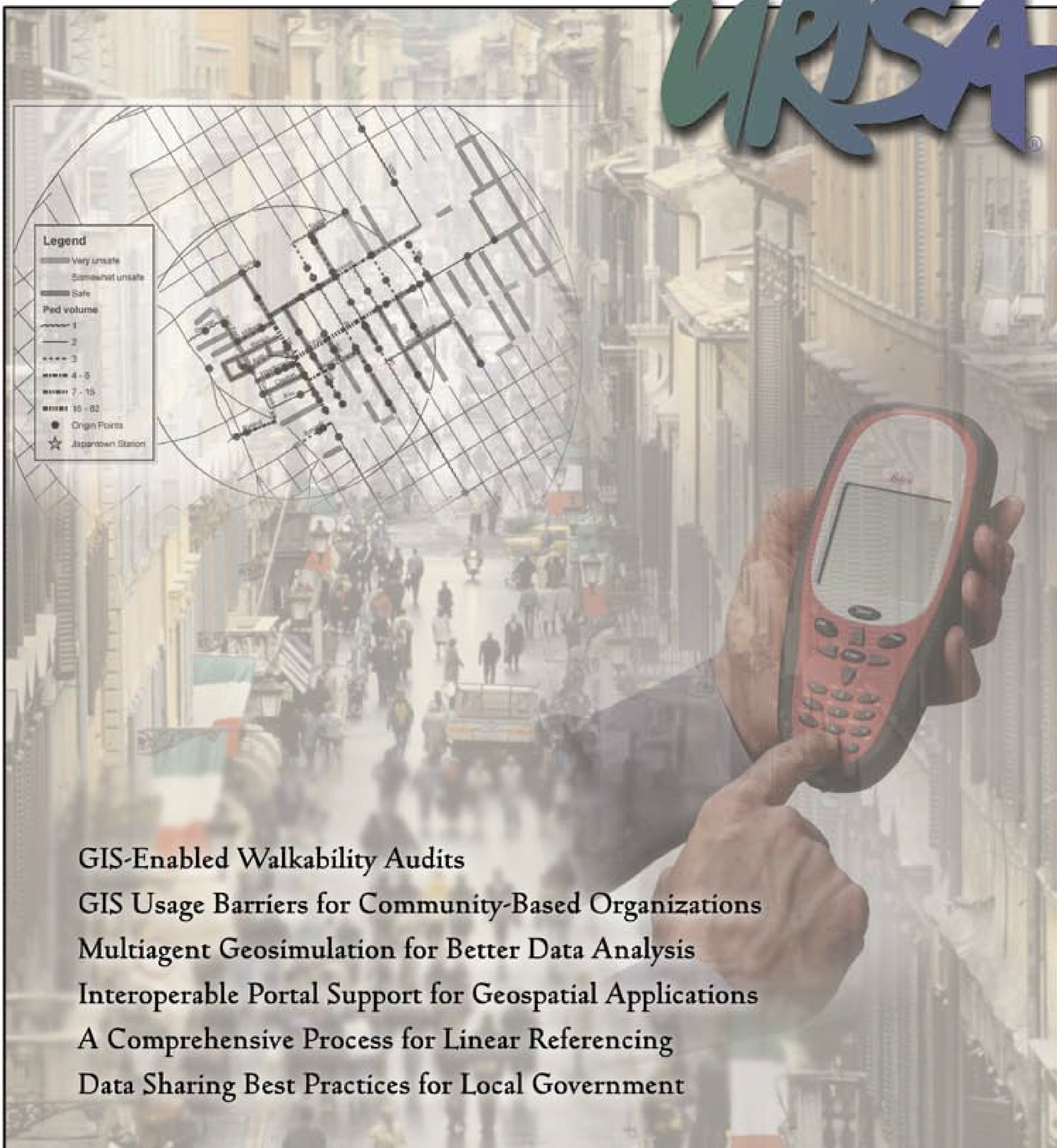
Journal

of the Urban and Regional Information Systems Association

Vol. 19 No. 2
2007



GIS-Enabled Walkability Audits

GIS Usage Barriers for Community-Based Organizations

Multiagent Geosimulation for Better Data Analysis

Interoperable Portal Support for Geospatial Applications

A Comprehensive Process for Linear Referencing

Data Sharing Best Practices for Local Government

Volume 19 • No. 2 • 2007

Journal of the Urban and Regional Information Systems Association

Contents

Refereed

On the Cover:

A common and humorous test of mental dexterity is the ability to walk and chew gum at the same time. These two simple tasks rarely conflict and are performed without thought. Over the past 10 years a new test of mental dexterity has arisen with vast implications in the field of geospatial technology; can one walk and collect data at the same time?

Research on walking and the built environment has accelerated by an increased interest in the relationship between urban form and public health. As the research has progressed, so has the interest in developing ways to collect data at a very fine scale—in essence, to be able to collect data at the streetscape level and link this data to transportation behavior.

An article by Marc Schlossberg, Asha Weinstein, and Katja Irvin entitled An Assessment of GIS-enabled Walkability Audits examines the development and implementation of a GIS-based pedestrian audit tool that allows users to collect data in electronic form using a handheld computer. The paper, which highlights Volume 19, Number 2 of the URISA Journal, identifies current shortcomings with walkability audit tools but concludes that appropriately applying GIS-enabled pedestrian audit tools can be an efficient way to collect and quickly analyze pedestrian infrastructure characteristics so that planners, practitioners, policy makers, and community members can make more effective decisions.

Journal

Publisher: **Urban and Regional Information Systems Association**

Editor-in-Chief: **Jochen Albrecht**

Journal Coordinator: **Scott A. Grams**

Electronic Journal: http://www.urisa.org/journal.htm

EDITORIAL OFFICE: Urban and Regional Information Systems Association, 1460 Renaissance Drive, Suite 305, Park Ridge, Illinois 60068-1348; Voice (847) 824-6300; Fax (847) 824-6363; E-mail info@urisa.org.

SUBMISSIONS: This publication accepts from authors an exclusive right of first publication to their article plus an accompanying grant of non-exclusive full rights. The publisher requires that full credit for first publication in the *URISA Journal* is provided in any subsequent electronic or print publications. For more information, the “Manuscript Submission Guidelines for Refereed Articles” is available on our website, www.urisa.org, or by calling (847) 824-6300.

SUBSCRIPTION AND ADVERTISING: All correspondence about advertising, subscriptions, and URISA memberships should be directed to: Urban and Regional Information Systems Association, 1460 Renaissance Dr., Suite 305, Park Ridge, Illinois, 60068-1348; Voice (847) 824-6300; Fax (847) 824-6363; E-mail info@urisa.org.

URISA Journal is published two times a year by the Urban and Regional Information Systems Association.

© 2007 by the Urban and Regional Information Systems Association. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by permission of the Urban and Regional Information Systems Association.

Educational programs planned and presented by URISA provide attendees with relevant and rewarding continuing education experience. However, neither the content (whether written or oral) of any course, seminar, or other presentation, nor the use of a specific product in conjunction therewith, nor the exhibition of any materials by any party coincident with the educational event, should be construed as indicating endorsement or approval of the views presented, the products used, or the materials exhibited by URISA, or by its committees, Special Interest Groups, Chapters, or other commissions.

SUBSCRIPTION RATE: One year: $295 business, libraries, government agencies, and public institutions. Individuals interested in subscriptions should contact URISA for membership information.

US ISSN 1045-8077

Editors and Review Board

URISA Journal Editor

Editor-in-Chief
Jochen Albrecht, *Department of Geography, Hunter College City University of New York*

Thematic Editors

Editor-Urban and Regional Information Science
Vacant

Editor-Applications Research
Lyna Wiggins, *Department of Planning, Rutgers University*

Editor-Social, Organizational, Legal, and Economic Sciences
Ian Masser, *Department of Urban Planning and Management, ITC (Netherlands)*

Editor-Geographic Information Science
Mark Harrower, *Department of Geography, University of Wisconsin Madison*

Editor-Information and Media Sciences
Michael Shiffer, *Department of Planning, Massachusetts Institute of Technology*

Editor-Spatial Data Acquisition and Integration
Gary Hunter, *Department of Geomatics, University of Melbourne (Australia)*

Editor-Geography, Cartography, and Cognitive Science
Vacant

Editor-Education
Karen Kemp, *Director, International Masters Program in GIS, University of Redlands*

Section Editors

Software Review Editor
Jay Lee, *Department of Geography, Kent State University*

Book Review Editor
David Tulloch, *Department of Landscape Architecture, Rutgers University*

Article Review Board

Peggy Agouris, *Department of Spatial Information Science and Engineering, University of Maine*

Grenville Barnes, *Geomatics Program, University of Florida*

Michael Batty, *Centre for Advanced Spatial Analysis, University College London (United Kingdom)*

Kate Beard, *Department of Spatial Information Science and Engineering, University of Maine*

Yvan Bédard, *Centre for Research in Geomatics, Laval University (Canada)*

Barbara P. Buttenfield, *Department of Geography, University of Colorado*

Keith C. Clarke, *Department of Geography, University of California-Santa Barbara*

David Coleman, *Department of Geodesy and Geomatics Engineering, University of New Brunswick (Canada)*

David J. Cowen, *Department of Geography, University of South Carolina*

Massimo Craglia, *Department of Town & Regional Planning, University of Sheffield (United Kingdom)*

William J. Craig, *Center for Urban and Regional Affairs, University of Minnesota*

Robert G. Cromley, *Department of Geography, University of Connecticut*

Kenneth J. Dueker, *Urban Studies and Planning, Portland State University*

Geoffrey Dutton, *Spatial Effects*

Max J. Egenhofer, *Department of Spatial Information Science and Engineering, University of Maine*

Manfred Ehlers, *Research Center for Geoinformatics and Remote Sensing, University of Osnabrueck (Germany)*

Manfred M. Fischer, *Economics, Geography & Geoinformatics, Vienna University of Economics and Business Administration (Austria)*

Myke Gluck, *Department of Math and Computer Science, Virginia Military Institute*

Michael Goodchild, *Department of Geography, University of California-Santa Barbara*

Michael Gould, *Department of Information Systems Universitat Jaume I (Spain)*

Daniel A. Griffith, *Department of Geography, Syracuse University*

Francis J. Harvey, *Department of Geography, University of Minnesota*

Kingsley E. Haynes, *Public Policy and Geography, George Mason University*

Eric J. Heikkila, *School of Policy, Planning, and Development, University of Southern California*

Stephen C. Hirtle, *Department of Information Science and Telecommunications, University of Pittsburgh*

Gary Jeffress, *Department of Geographical Information Science, Texas A&M University-Corpus Christi*

Richard E. Klosterman, *Department of Geography and Planning, University of Akron*

Robert Laurini, *Claude Bernard University of Lyon (France)*

Thomas M. Lillesand, *Environmental Remote Sensing Center, University of Wisconsin-Madison*

Paul Longley, *Centre for Advanced Spatial Analysis, University College, London (United Kingdom)*

Xavier R. Lopez, *Oracle Corporation*

David Maguire, *Environmental Systems Research Institute*

Harvey J. Miller, *Department of Geography, University of Utah*

Zorica Nedovic-Budic, *Department of Urban and Regional Planning,University of Illinois-Champaign/Urbana*

Atsuyuki Okabe, *Department of Urban Engineering, University of Tokyo (Japan)*

Harlan Onsrud, *Spatial Information Science and Engineering, University of Maine*

Jeffrey K. Pinto, *School of Business, Penn State Erie*

Gerard Rushton, *Department of Geography, University of Iowa*

Jie Shan, *School of Civil Engineering, Purdue University*

Bruce D. Spear, *Federal Highway Administration*

Jonathan Sperling, *Policy Development & Research, U.S. Department of Housing and Urban Development*

David J. Unwin, *School of Geography, Birkbeck College, London (United Kingdom)*

Stephen J. Ventura, *Department of Environmental Studies and Soil Science, University of Wisconsin-Madison*

Nancy von Meyer, *Fairview Industries*

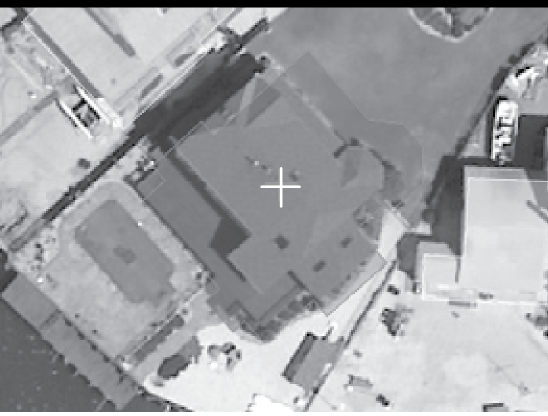
Barry Wellar, *Department of Geography, University of Ottawa (Canada)*

Michael F. Worboys, *Department of Computer Science, Keele University (United Kingdom)*

F. Benjamin Zhan, *Department of Geography, Texas State University-San Marcos*

Introducing **Change*Findr*™**

*The World's Most Advanced **Patented** Technology For Change Detection And Change Analysis**


Existing Ortho before change


Pictometry Oblique before change


Pictometry Oblique after change

Reveal. **Review. Recoup.**

ChangeFindr™ is the combination of **Pictometry's Change Analysis™ software** and **Hitachi's HouseDiff® service** that automatically outlines and identifies changes in any structure when comparing two image data sets.

This exciting solution enables you to review change over time intervals that you select. This allows you to capture unreported or under-reported improvements quickly and easily during your discovery process. **ChangeFindr** gives you the best of both worlds: verify reported changes yourself or use **Pictometry's Technical Services** to provide you with the verification report.

The benefits? Find hidden revenues, reduce expenses, improve productivity, as well as increase your tax base year after year, all from the convenience of your desktop PC.

How Pictometry and HouseDiff® can help:

*"By using both Pictometry and HouseDiff® we have found over **one-and-a-half million dollars** in new valuations."*

What tax assessment professionals are saying about Pictometry:

*"Our Appraisal Staff and Planning Commission GIS users have dramatically reduced their number of site visits**, in some cases by 70%,** by using Pictometry."*

*"In less than a year of actual use, we have **recovered more costs than what it cost us."***

Let Pictometry's innovative technology expand your productivity as it has for hundreds of assessors. To learn more visit www.pictometry.com/assessor

Pictometry®
Intelligent Images®

www.pictometry.com
888-771-9714

* Pictometry and Hitachi HouseDiff® each have patented components used in ChangeFindr. HouseDiff® is a registered service mark of Hitachi Software Engineering Co., Ltd.

FUTUREview 2008
BEYOND THE IMAGE

SAVE THE DATE
OCTOBER 27-29
SAN ANTONIO • TX

An Assessment of GIS-Enabled Walkability Audits

Marc Schlossberg, Asha Weinstein Agrawal, and Katja Irvin

Abstract: *Research on walking and the built environment is a fairly recent area of inquiry, accelerated over the past ten years by an increased interest in the relationship between urban form and public health. As the research has progressed, so has the interest in developing ways to collect data at a very fine scale—in essence, to be able to collect data at the streetscape level and link this data to transportation behavior. This paper discusses the development and implementation of a GIS-based pedestrian audit tool that allows users to collect data in electronic form using a handheld computer (i.e., a pocket PC or personal digital assistant (PDA)). While, such tools may be useful for better understanding the relationship between the built environment and pedestrian behavior, the tools may be unnecessarily complex and unfocused. Specifically, these walkability audit tools could be improved by: (1) applying unique sets of walkability measures to different types of walking environments; (2) perhaps focusing auditing activities on major streets and intersections only (e.g., do not audit neighborhood streets where possible); (3) including subjective as well as objective measures of the streetscape; (4) verify the accuracy of digital base maps before widespread implementation; and (5) continuously evaluating whether the simpler technology of pen and paper would be preferable alternatives. The authors conclude that appropriately applying GIS-enabled pedestrian audit tools can be an efficient way to collect and quickly analyze pedestrian infrastructure characteristics so that planners, practitioners, policy makers, and community members can make more effective decisions on behalf of walkability.*

INTRODUCTION

Research on walking and the built environment is a fairly recent area of inquiry, accelerated over the past ten years by an increased interest in the relationship between urban form and public health. As the research has progressed, so has the interest in developing ways to collect data at a very fine scale—in essence, to be able to collect data at the streetscape level and link this data to active transportation behavior (Schlossberg 2007). However, the lack of quick and cost-effective tools for collecting block-by-block data about the streetscape has prevented more widespread research and application of such tools. This paper discusses the development and implementation of a GIS-based pedestrian audit tool that allows users to collect data in electronic form using a handheld computer (i.e., a pocket PC or personal digital assistant (PDA)).

In a recent article on visualizing and measuring walkability, Schlossberg suggested looking forward to new technological approaches. In theory, tools that could allow for detailed, GIS-enabled data collection about pedestrian environments on handheld computers would allow planners to better understand the relationship between specific characteristics of the built environment and their relationship to either overall walking within an area or preferences for walking along one route or another. Once this relationship between the walking environment and walking behavior could be established, then specific recommendations to improve walking conditions could be made to policy makers, planners, transportation officials, and other decision makers could be made to improve conditions for walking.

This paper discusses the development and implementation of such a pedestrian audit tool and evaluates its promise for use in future projects. The larger project within which this tool was developed examined correlations between aspects of the built environment with the actual route choices that people make when walking from home to their nearest transit stations. The variables within this GIS-based audit tool were adapted from existing research literature and from other pedestrian audit tools, namely the Pedestrian Environment Data Scan (PEDS) developed by Dr. Kelly Clifton and Andrea Livi at the University of Maryland and Dr. Daniel Rodriguez at the University of North Carolina. This paper, however, will not focus on the types of data collected; rather, what follows is an evaluation of using this technology as a way to gather more detailed and nuanced data about walkable environments.

CONCEPTUAL BACKGROUND

As researchers, practitioners, and policy makers have become increasingly interested in the relationship between the built environment and physical activity, recognition of tools that appropriately measure urban form at a pedestrian scale has also increased. Perhaps the best-known and utilized tool in this area is an environmental audit instrument called SPACES, a comprehensive tool that helps inventory the characteristics of and along a roadway segment (Pikora, Giles-Corti et al. 2003). The authors categorize different factors of a walking environment into five classifications: (1) functional (physical attributes of the street), (2) safety (characteristics of a safe environment), (3) aesthetic (elements such as trees or gardens), (4) destination (relationship of neighborhood services to residences), and (5) subjective.

Moudon and Lee (2003) focused their work in a similar area, but dedicate more time to developing a conceptual framework for measuring walkability to help direct future research efforts. To develop their framework, they performed an exhaustive review of more than 30 published methodologies and inventorying tools that have been developed to assess walkability. They outlined a theoretical framework called the Behavioral Model of Environments (BME) that seeks to account for personal, physical, and internal responses factors that may explain the connection between individual pedestrians and their walking environments. In essence, they are attempting to lay the theoretical framework for focusing on the characteristics of place when considering the relationship between urban form and pedestrian behavior.

McMillan (2005) also provides a conceptual framework for others to follow when studying urban form and pedestrian accessibility. Through her focus on a child's trip to school, McMillan realized that urban form does not directly impact how a child gets to school, but rather a set of mediating and moderating factors do (McMillan 2005). Mediating factors include neighborhood and traffic safety, as well as household characteristics such as the availability of automobiles at home and the distance between home and school. Moderating factors include parental attitudes, social or cultural norms, and sociodemographic characteristics (McMillan 2005). The importance of this research is to illustrate that while urban form may have an important impact on pedestrian behavior, other nonbuilt environmental factors are also important. And in terms of our interest in audit tools themselves, McMillan's research shows that such tools may yield important data about the walking environment, but that such urban form data only represents one component of the overall decision-making process to walk or not.

Clifton and Livi (2005) developed the Pedestrian Environment Data Scan (PEDS) audit tool, which included 78 measures of streetscape characteristics that other research has shown to influence walkability. In addition to developing their tool, Clifton and Livi studied the inter-rater reliability of the instrument to understand the potential of such tools to be used in broad geographic areas with a diversity of audit administrators. Despite a wide range of street segment uses, conditions, and aesthetics, they found relatively high reliability scores for many of the questions contained within the audit instrument. Ewing (2006) utilized input from urban design professionals to develop operational definitions of the built environment relevant to pedestrians and translated those definitions into a field survey instrument (Ewing, Handy et al. 2006).

While the development of these conceptual and operational frameworks for assessing local walkability are important, researchers have been limited by the amount of time required to conduct block-by-block assessments of every street segment and intersection within a study area. As researchers identify more aspects of the built environment that may be important in creating walkable environments, the burden of applying those measures to each street segment grows. In their study of two pedestrian and bicycling environmental audit tools, Emery found it challenging to reliably evaluate road segments (Emery, Crump et al. 2003).

Recent technological innovations, however, have made it possible to create audit tools that work within a GIS environment on PDAs, enabling data collection that is relatively quick and instantly embedded into an electronic and GIS-ready data format for more prompt analysis. We used just such a tool in our study of pedestrian route choice, and the remainder of this paper will report on this method of data collection.

PROJECT AND AUDIT TOOL BACKGROUND

We developed the pedestrian audit tool as part of a larger research project designed to understand how far people walk to rail stations in the morning and what routes they take. Stages of the project included surveying people who walked to transit stations, asking them about their reasons for choosing the routes they do, mapping the routes they took to the station the day of the survey, and then auditing the physical environment to determine whether the built environment impacted the routes pedestrians took to access transit.[1] Our goal was to collect data about the built environment for every street segment and intersection within one mile of four transit stops. We used the PEDS tool as our starting point for developing the audit tool, but heavily modified the tool by adding and deleting variables to make the tool relevant to our particular study. Once we decided on the appropriate variables to use, we then created an interface to allow the tool to work with GIS software on a PDA.[2] We felt it was important to have different variables to audit the street segments and to evaluate intersections, so we created separate sets of variables to do just that. We field-tested the tool and made modifications before collecting data.

In all, our street segment audit consisted of six screen pages per street segment, constituting approximately 25 different sets of variables, depending on the presence or absence of certain features. The intersection audit consisted of two screen pages and six variables (see Figure 1 for sample screenshots of the tool).

We designed the audit tool to make it easy to use. In particular, we minimized the amount of screen "typing" and created data-entry forms that were efficient and intuitive for the auditor. To do so, we used a variety of methods for data input, including checkboxes, drop-down menus, and text fields where we could enter in observations that deviated from our predefined answer responses. In addition to these standard questions, we customized the data-input interface with more advanced features designed to make data collection more efficient. One feature was the use of skip sequences to make the flow through the tool more seamless, and another was the use of an on-the-fly field default-setting feature. In this default-setting feature, the user could save the answers entered for one particular street segment and then load them automatically for the next street segment, thus minimizing the data to be entered for the second segment. In general, street segments do not differ much from one block to the next, so this automatic field completion feature was a useful time-saving addition. It was still possible to alter answers to specific questions in the default setting as needed, when one or two features of a block segment were unique.

Collecting data with this tool essentially involved walking

The screenshots on the top row represent three of the six data entry screens for assessing street segments. To the right is one of the two screens for assessing intersections. To the far right is a screenshot of the GIS software loaded onto a PDA. The data entry screens appear on the PDA when a street segment or intersection is tapped by the user for data entry.

Figure 1. Examples of the walkability audit data-entry forms

along each street segment (or intersection) and visually scanning the environment as the audit questions were answered. At first this involved walking most of a street segment, pausing to look around, and then entering data while conducting additional visual scans if necessary. Over time, as the auditors became more familiar with the local environment and comfortable with the audit tool, the data-collection process became more efficient; essentially the auditors were able to walk, scan, and enter data simultaneously. Thus, in some environments there was little additional time needed for data collection other than the time it takes to walk along each street segment.[3] That said, there is an incentive to minimize the number of audit questions because the more questions for any particular segment, the greater the concentration needs for the auditor and the greater the likelihood of error. The default function mentioned previously helped significantly to achieve this efficiency without compromising quality.

To enter data, the PDA needs to be preloaded with a map of the local area as well as the database entry tool. In our case, we loaded the PDAs with edited TIGER street maps and intersection point maps that we created separately with GIS software. The auditor located the street segment on the PDA map where he or she was standing and simply tapped the street segment (or intersection) on the screen to call up the data-entry forms. Once the questions were answered, the data was automatically saved and associated with the street segment in the GIS file; thus, only one data-entry step is needed to both collect the audit data and prepare it for GIS analysis. This process was repeated until the audit of the study area was completed.

Overall completion time depended both on the size of the study area and the quantity of road segments in the study area. In our study, the areas audited in the Portland, Oregon, region had much higher densities of streets around transit stops than the areas around the other study areas in San Jose and Oakland, California. Budgeting time for conducting the audit, therefore, depends on the number of segments and intersections within a study area, and not just on the size of the study area itself. Total data-collection

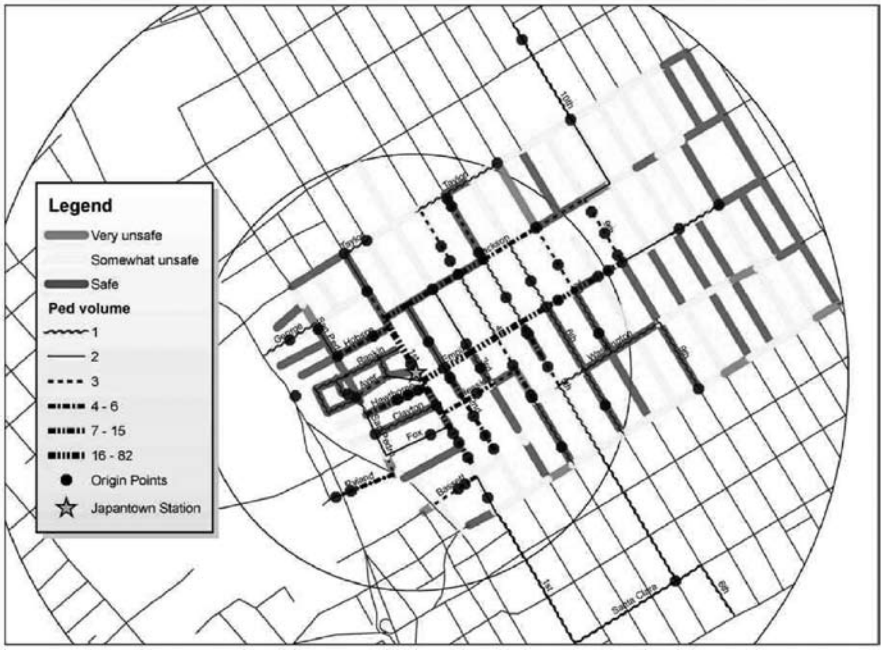


Figure 2. Map of safety overlaid with actual pedestrian routes to transit

dataset that would otherwise be impossible to derive. Before using the tool in the field, however, it is critical to ground truth the street base map that will form the core of the dataset. As mentioned previously, we utilized the TIGER street file as our base map and we did so because we wanted to use a freely accessible source of data that would be available to any community in the United States for free. As is often the case with TIGER data, some errors exist between the TIGER map and the actual presence or absence of streets around the transit stops. In some cases, the TIGER data included streets that did not exist, and in others streets existed that were not part of the TIGER data. And TIGER data is not particularly good at including off-street paths.

It is also important to check the address ranges of the streets within the TIGER data to ensure they are consistent with actual address ranges of the streets. We have experienced address ranges that were one block off, meaning we had to correct these errors in the map by hand before it was possible to accurately geocode our survey data. It is possible to add or delete street segments or adjust street ranges in the field on the PDA, but it is critically important that some basic ground truthing of the base GIS data be conducted prior to auditing the environment. Alternatively, one could use locally produced base map files that would presumably be more up-to-date and accurate for local conditions.

time, thus, is a function of the average length of a street segment, the total number of street segments, the total data-entry questions per street segment, the total numbers of intersections, and the total number of audit questions for those intersections.

Once the data is collected, it can be mapped easily because the data was entered directly into a GIS format. Figure 2 shows an example of a map created by using a pedestrian audit tool.

REFLECTIONS AND MODIFICATIONS

Based on our experience in developing pedestrian audit instruments for classroom and research projects, and with the recent completion of a major study of transit access by pedestrians, we offer five general areas for consideration when developing pedestrian audit tools:

1. Verify the accuracy of the street base map before recording any data.
2. Develop data-gathering tools customized to street/path type.
3. Consider focusing data collection only on major streets.
4. Determine the relative importance of subjective versus objective measures.
5. Evaluate the utility of digital versus paper-and-pen technology for conducting audits.

GROUND TRUTH BASE MAPS

In our experience, with an accurate base map of streets (or sidewalks), collecting GIS-enabled data at a streetscape level was generally straightforward and provided the analysis with a rich

Customizing Data Collection by Street Type

Based on our experience, we have concluded that when auditing pedestrian environments, it would be useful to be able to differentiate between street type when collecting data and then collect data unique to each type of street or path. It became clear after collecting data that arterial and collector streets required a different set of attributes to collect compared to neighborhood streets. For example, street width, sidewalk buffers, on-street parking, and the number of high-volume driveways to cross were all much more important on arterials and collectors where the volume and speed of vehicles presents much more of a safety threat and level of discomfort compared to neighborhood streets. On neighborhood streets, at least in our study areas, aesthetic and street features seemed to be less influential for a pedestrian's ability to walk. For these streets, perhaps the most important feature would be the presence of sidewalks. Customizing data-entry variables for different types of streets would streamline the data-collection process and allow a greater range of streets to be surveyed in a shorter period of time. As mentioned previously, the tool we used had a built-in feature to save previous data-entry

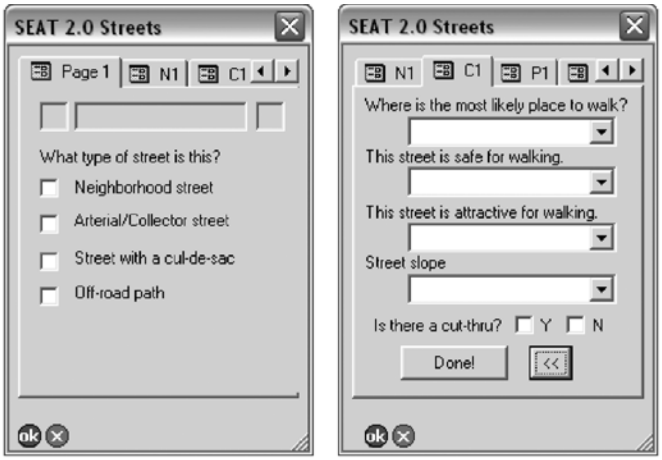


Figure 3. An example of an audit tool that is customized by street type

responses in a default mode so that they could be recalled for a subsequent street segment, but this feature required additional computer programming expertise that could well be avoided with a revised approach to data gathering that customizes the data fields by street type.

Figure 3 shows an example of a potential data-filtering system by street type. These are two screenshots from a new tool, the School Environment Assessment Tool (SEAT), being developed to audit walkability for safe routes to school. The image on the left is the initial data-entry page that appears, which provides an initial filter as to the street type being audited. Subsequent pages are customized based on which street type is selected. The image on the right is the data-entry screen that appears for a street segment that ends with a cul-de-sac. Most streets that end in cul-de-sacs are neighborhood roads with low volumes of cars and are most likely not severely impacted by different measures of walkability. Documenting whether a pedestrian can cut through the end of the cul-de-sac, however, is important and because it only pertains to segments ending in cul-de-sacs, this question only appears for streets selected as cul-de-sacs on the first data-entry page.

FOCUS ON ARTERIALS/ COLLECTORS

It was also apparent from our study sites that in some study areas, it was almost unnecessary to audit neighborhood streets and that focusing the audit on arterials, collectors, and their associated intersections may have been a better use of data-gathering time. In some neighborhoods, all neighborhood streets had sidewalks and were pleasant and safe enough to walk along. The real key, though, for evaluating the potential pedestrian friendliness of one's journey from home to transit (or other destination) were the attributes of the major roads and the intersections between neighborhood roads and major roads. In essence, the more focused question could be: What makes a major automobile road more or less pedestrian-friendly? In this approach, all neighborhood streets could be assumed to be generally walkable and the focus would concentrate on making the areas where high volumes and speeds of automobiles and higher volumes of pedestrians would intersect safer and more enjoyable for pedestrians to navigate. It is in these high-intersection places where interventions on behalf of walking might be best targeted. Comparing route choices and route avoidance by pedestrians along these more major streets would allow planners and policy makers to better focus resources and interventions where they would be most needed, and the audit data could point these decision makers into appropriate directions for their interventions. Of course, in study areas where sidewalks are not universally present, or where street widths in particular vary quite a bit, and are deemed important barriers for walking, then including neighborhood roads in the audit may be important.

Comparing Subjective to Objective Measures

An area for further investigation is to measure the relationship between broad subjective evaluations, such as "Is this street attractive and safe for walking?" with the objective measurements of individual streetscape traits. In other words, can a simple subjective question such as this be enough to evaluate streets and intersections on walkability principles, even with the eventual variation in subjective evaluations? Certain street segments in our audit evaluation felt like poor environments to walk along because of aesthetics, proximity to heavy traffic, and a general feeling of being uncomfortable places. It would be easy to imagine that pedestrians would simply choose parallel paths to walk along. Using the specific measures contained in the audit tool, however, the general feeling was not conveyed. For example, one of these uncomfortable walking streets had a buffer with trees between the sidewalk and the street, on-street parking, only two travel lanes in each direction, and properties that were decently maintained, among other attributes that would supposedly contribute to a safe and attractive walking environment. But by focusing on the individual, objectively measuring details of the streetscape, the larger feeling of the area could be missed without the inclusion of the subjective evaluations. Perhaps the individual objective measures we used were incomplete in this regard, or perhaps no matter how hard we try to measure every aspect of the built environment objectively, it is our subjective perception of a place that is the most insightful evaluation. This is the conclusion reached in a recent evaluation of audit tools where the authors conclude that "walking behavior is better explained by perceptions than sociodemographics or objective assessment of the environment" (Livi Smith and Clifton 2006). Of course, relying only on subjective evaluations does not provide decision makers with any guidance on how to design or retrofit good walkable locations; but for studies of pedestrian route preference, such questions may be enough to determine whether urban design has impact on route choices or not, or whether shortest routes is the predominant factor in influencing trip making.

Paper Versus PDA

This last area centers on the utility of an electronic and GIS-enabled approach to audit data gathering versus a more traditional approach of paper, pen, and clipboard. The obvious benefit of the handheld GIS computer approach is that by collecting data both in an electronic and a GIS format, there is no need for subsequent data entry once the audit is complete. When entering data collected by paper or data collected electronically with a handheld database program, there is additional risk of data-entry error when converting the data to a GIS environment. With handheld GIS technology, that risk is minimized because data can be collected in closed-ended questions directly within a GIS environment. In addition to potential error of additional data entry, there can be a significant amount of time involvement, especially with paper-based audit tools, in converting them to a usable digital format.

The handheld computer approach has the additional benefit of instant mapmaking, which may be important for community-based approaches to walkability assessments. For example, a group of community or elementary school volunteers may use the audit tool to assess streets and intersections within a mile of a target school. With the handheld GIS approach to conducting pedestrian audits, it would be possible for this group of volunteers to easily collect data in a few hours, gather together at the end of data collection, and synthesize the data from each PDA used into a single data file that can be mapped on the spot. Incorporating portable printer technology would allow each volunteer to leave the day's auditing with initial walkability maps based on data collected that day. For community-based approaches to walking issues, the ability to transform volunteer energy into a tangible map can be vital in sustaining community interest and catalyzing decision makers into taking appropriate action in regards to the needs of pedestrians.

Of course, the use of this advanced technology in assessing the walking environment can be limiting or carry risks. Perhaps the biggest limitation of handheld computer technology is taking field notes. When conducting a walkability audit, sometimes making field notes for specific audit questions is desirable, and unless specifically programmed to do so, such open-ended note taking can be more limiting on the handheld computers. Technologically, the built-in word-processing, voice-recording, or picture-taking capabilities of PDAs can be used, but writing observations or comments directly onto a survey form is probably still easier to do with pen and paper.

Another limitation of the digital approach is that audit questions are permanently preordered and auditors are forced to answer audit questions as they are written, not as they are observed. Paper versions of audits allow the auditor to answer questions in the most logical order for what is being observed, but electronic approaches make this approach too cumbersome to be useful.

Other technology issues are that the battery lives of PDAs can be short for all-day auditing, unless extended batteries are purchased. Also, in our experience, some people just find that PDAs are too cumbersome to use and that paper and pen are much simpler and easy to understand than using GIS software on a handheld computer. Good training and preparation can overcome this hurdle, however. Finally, carrying expensive computers while analyzing neighborhood streets and sidewalks can be unsafe in certain neighborhoods (or make auditors feel unsafe), especially if auditing teams are perceived as outsiders to that neighborhood. Making good community connections, as should be done with any project where a potential problem of outsider versus insider may exist, should be a prerequisite to conducting the auditing work.

CONCLUSION

An earlier study by Schlossbert (2007) found that "[Pedestrian audit tools] allow planners to begin to collect the more nuanced characteristics of an area that makes it more or less attractive for pedestrians. And instead of classifying streets based on automobile-oriented categories like minor, arterial, or collector road, with [instruments like these], streets can be classified with a pedestrian-orientation. GIS analysis can then distinguish between paths based on more relevant variables" (Schlossberg 2007). Indeed, walkability audit tools hold much promise to allow researchers and practitioners to better understand the relationship between elements of the built environment and pedestrian behavior. Most pedestrian trips are short, so it makes sense that tools are emerging that can be applied to small-area analyses. We have found that these audit tools, especially when used within a GIS environment, generally provide efficient ways for collecting data and provide researchers with data that can be more quickly analyzed with less error than similar data collected with pen and paper.

Because the increasing efficiency of these tools makes it tempting to collect more data, this urge should be resisted, or at least carefully considered. The appropriate use of this technology is important, and in conducting walkability audits, an appropriate application may be one that focuses on the real areas of pedestrian conflict, rather than on all streets and all elements of urban form. We believe that as these tools evolve, they can be used appropriately to understand how pedestrians interact with their local environment and that these audit instruments will be one means for enhancing the pedestrian infrastructure in such a way as to support increased rates of walking in our communities.

Acknowledgments

We would like to acknowledge and thank the Mineta Transportation Institute and the Oregon Transportation Research and Education Consortium (OTREC) for their support of this work.

About the Authors

Marc Schlossberg is an associate professor in the Planning, Public Policy, and Management Department at the University of Oregon, Eugene, an associate director of the Oregon Transportation Research and Education Consortium (OTREC), and a research associate of the Mineta Transportation Institute. He teaches and researches issues on participatory GIS, active modes of transportation, and urban form.

Corresponding Address:
Planning, Public Policy, and Management Department
University of Oregon
128 Hendricks Hall
Eugene, OR 97403
541-346-2046
Fax: 541-346-2040
schlossb@uoregon.edu

Asha Weinstein Agrawal is an assistant professor in the Department of Urban and Regional Planning at San José State University and a research associate of the Mineta Transportation Institute. She teaches and researches in the areas of transportation planning and finance, and the history of pedestrianism.

Katja Irvin is a planner with PMC in Rancho Cordova, Calfornia. She recently received an M.V.P. degree from San José State University, where she was a research assistant for the Mineta Transportation Institute.

References

Emery, J., C. Crump, et al. 2003. Building the methods—reliability and validity of two instruments designed to assess the walking and bicycling suitability of sidewalks and roads. American Journal of Health Promotion 18(1): 9.

Ewing, R., S. Handy, et al. 2006. Identifying and measuring urban design qualities related to walkability. TRB 85th Annual Meeting Compendium of Papers CD-ROM, Washington, D.C., Transportation Research Board.

Livi Smith, A., and K. Clifton. 2006. The relationship between neighborhood environment and walking behavior: the mediating influence of perceptions. 11th International Conference on Travel Behaviour Research, Kyoto, Japan.

McMillan, T. 2005. Urban form and a children's trip to school: the current literature and a framework for future research. Journal of Planning Literature 19(4): 440–56.

Pikora, T., B. Giles-Corti, et al. 2003. Developing a framework for assessment of the environmental determinants of walking and cycling. Social Science and Medicine 56(8): 12.

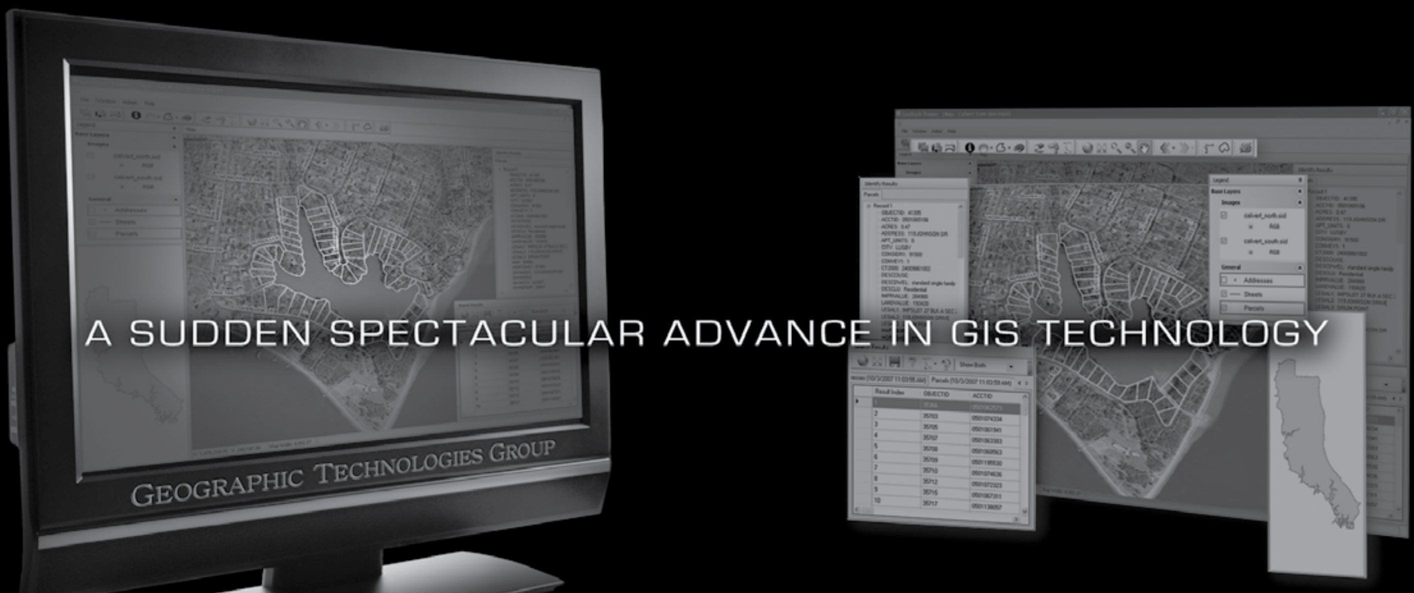
Schlossberg, M. 2007. From TIGER to audit instruments: using GIS-based street data to measure neighborhood walkability. Transportation Research Record: Journal of the Transportation Research Board 1982: 48-56.

Footnotes

1 The research project was sponsored by the Mineta Transportation Institute, and the full project report will be available on its Web site at http://transweb.sjsu.edu/.

2 We used ArcPad GIS software created by ESRI on Dell Axim PDAs.

3 It is likely that one's pace is a bit slower when entering data compared to someone who would be walking the same street segment without a data-entry task to complete, but over time this tool does allow the auditor to essentially collect data without breaking a stride.



www.geotg.com/geoblade

Built on ESRI's ArcGIS Sever® technology Geographic Technologies Group® introduces the most advanced and truly enterprise GIS solution on the market. A solid, reliable, and intuitive interface that guarantees a Return on Investment (ROI). Designed and engineered to offer local government professionals a "core" GIS tool, GTG's GeoBlade software offers an interface and GIS functionality that is "second to none". This revolutionary product consists of a host of custom department specific "GeoBlades" or interfaces. A truly spectacular solution!

GeoBlade™

Geographic Technologies Group®

1.888.757.4222 • www.geotg.com

Institutional and Organizational Barriers to Effective Use of GIS by Community-Based Organizations

Ann-Margaret Esnard

Abstract: *This research was sparked by a general observation that community-based organizations (CBOs) do not use GIS effectively for community and land-redevelopment initiatives even when software, hardware, training, data, formal partnership agreements, and access to GIS services are all in place. A sample pool of CBOs was surveyed to gather information about their community planning and land-(re)development work, staffing, missions, geographic boundaries, GIS capacity, and perceived barriers to the effective use of GIS. Several findings and barriers are reported in the paper, including lack of a mission statement that promotes the use of GIS and related technologies; difficulty in applying technology to existing projects; the mismatch between actual and known applications of GIS; and the ability of CBOs to survive without GIS. A revised GIS implementation framework focused on organizational capacity, context, and analysis, instead of on the traditional "hardware, software, data, and people" components, is proposed in the final section of this paper.*

INTRODUCTION

Communities across the United States are undergoing rapid changes in land use, demographic, socioeconomic, and infrastructure conditions. Identification, visualization, monitoring, and evaluation of these changes and trends can be enhanced using geographic information systems (GIS) and online data and mapping technologies. GIS and related technologies can also be used to inform and assess land-(re)development projects and to develop comprehensive strategies in community redevelopment and affordable housing. Community-based organizations (CBOs), community development corporations (CDCs), neighborhood associations (NAs), and related coalitions (collectively referred to as community-based organizations in this paper), can take advantage of these technologies to mobilize assets that benefit communities (Greenstein 2006) and to (re)-create quality-of-life communities and regions.

In an ideal setting, these organizations can work with the same data and technological capabilities that many public and private agencies use. National examples and case studies that highlight how community-based organizations have used GIS for community change have been published (Boyd and Chan 2002, LISC and PolicyLink 2000); participatory planning processes and public participation GIS have served as a framework for many projects (Talen 2000, Al-Kodmany 2000, Craig et al. 2002); national online resources such as Data Place (http://www.dataplace.org) and Social Explorer (http://www.socialexplorer.com) are becoming commonplace; Web-GIS applications such as Neighborhood Knowledge California are being used in novel and unanticipated ways by community users (Rattray 2006); and there are greater efforts on the part of cities to make parcel data more accessible to the public via integrated parcel-based information systems (Chandler et al. 2006).

In June of 2003, the Lincoln Institute of Land Policy held a "GIS for CBOs" planning meeting that sparked the development of a workshop on that topic. As of 2007, the institute has held four such workshops: one in Cambridge, Massachusetts; two in Tampa, Florida; and one in Atlanta, Georgia (see the Appendix). To help community-based organizations keep up with rapidly changing technology while maintaining their mission, the customized workshops have focused on providing them with information about the functions and applications of GIS, general strategies for successful CBO-university-city government collaborative projects, capacity building and GIS implementation, and case studies on the types of land-development projects and analyses that can be enhanced with mapping and GIS analysis.

It is not clear, however, that community-based organizations use GIS effectively for community and land-redevelopment initiatives even when software, hardware, training, data, formal partnership agreements, and access to GIS services are all in place. Overall, there seems to be a general disconnect between the ease of access to data repositories and GIS/Web-GIS applications and the effective use of the technology for community planning and development. The assessment that community-based organizations have made great strides toward revitalizing distressed communities but have given insufficient attention to organizational capacity building and good organizational practices (Rodriguez and Herzog 2004, Glickman et al. 2004) might hold some clues to fleshing out this potential disconnect between access and effectiveness.

This paper begins with a condensed review of literature on institutional and organizational barriers and constraints that community-based organizations generally face, as well as background information about the survey sample pool. This is followed by findings on organizations' staffing, day-to-day activities, geographic scopes, GIS capacity, GIS application, and perceived barriers to the effective use of GIS. Concluding thoughts, as well as implications for professional development and related capacity-building initiatives, are offered in the final section.

GIS ADOPTION AND IMPLEMENTATION: OPPORTUNITIES AND CONSTRAINTS

Implementation of GIS technology in organizations is certainly not a new topic. Scholars such as Innes and Simpson (1993), Ventura (1995), and Nedovic and Pinto (1999) have documented barriers to GIS adoption in local government, the importance of the adaptability of these agencies, and the compatibility of the technology with existing organizational structures and cultures. This is supported by Sieber's research and findings on GIS implementation in grassroots conservation organizations, one of the most thorough discussions on institutional, organizational, technical, and political factors that limit such organizations in attracting and retaining GIS technology. According to Sieber (2000, 16), "simply laying an innovation on top of old processes will not induce the implementation to succeed."

Research has also shown that no perfect implementation model exists. Huxhold and Levinsohn (1995) provided a useful paradigm for planning, designing, and implementing GIS in municipal government and the private sector from a management and organizational perspective. Other scholars have suggested that local government organizations eventually adopt a mixed-model approach (Ramasubramanian 1999) rather a strict traditional top-down or bottom-up approach (Talen 2000) when implementing GIS. The same can be said for community-based organizations, although their capacity (or lack of) to adopt and adapt to technological changes and innovations is an important difference. Innes and Simpson's observability argument—i.e., "because benefits of GIS are difficult to observe and materialize in the distant future, they are unlikely to persuade decision-makers to fund GIS. . ." (Innes and Simpson 1995, 232)—is particularly pertinent to community-based organizations faced with major funding challenges and competing project and staffing priorities.

Characteristics such as institutional capacity, organizational capacity, and capacity building, typically discussed in an international context with an emphasis on community, regional and national health, and natural resource management concerns, might also explain institutional and organizational barriers to the effective use of GIS. Related factors include governance, leadership, missions, institutional inertia, resources (financial, technology, and human), tenure/maturity of the organization, services offered, communication, attitudes toward embracing new technology or ideas, and interorganizational and funding competition for limited resources (Ramasubramanian 1999, PathFinder International undated, Ramos 2001, and Roman and Moore 2004). Following is more detail on mission, human resources, technological resources, and leadership.

Enabling organizations to use and apply GIS technology in ways that enhance their work processes and fit into organizations' missions, scopes, and activities has been promoted by several scholars (Elwood and Leitner 1998a, Elwood and Leitner 1998b, Somers 1998). One purpose of a mission statement is its applicability as a planning/visionary tool (Roman and Moore 2004). Equitable and sustainable community planning and land redevelopment is a popular vision, but is oftentimes one of several activities. Some CBOs, for example, address community quality-of-life issues that range from workforce development skills to environmental justice (Esnard et al. 2004) in an attempt to serve the broad needs of diverse populations. Furthermore, CBOs affiliated with a consortium focus on neighborhoods that are at various stages of redevelopment, have different local geographic foci (e.g., sites/housing, main streets, neighborhood, region), and are grappling with specific local efforts (sometimes housing, transportation, school quality, or environmental quality) in relation to regional growth and development trends. There is often an unclear linkage between how technological resources and GIS-based analyses fit, if at all, into day-to-day activities and missions of such CBOs and consortia.

Human resources include paid staff, consultants, and volunteers. GIS implementation and capacity-building strategies with requirements for a dedicated GIS staff person is infeasible for most community-based organizations. Such organizations have had to rely on volunteers, student interns, and consultants to fulfill GIS needs, not a favorable situation for institutional stability and greater effectiveness, often attributed to full-time paid staff (Roman and Moore 2004). Salary levels, career advancement, and on-the-job education and training are other important factors for individuals considering GIS and other technological positions or roles in community-based organizations. Overall, specialized GIS jobs do not fit the typical organizational model and staff culture, where broad expertise is generally preferred.

Technological resources help increase an organization's overall outreach, communication, and workflow. Most community-based organizations have some computing infrastructure in place (word processing, e-mail, membership databases, and Web sites). However, they are at various stages with regard to GIS adoption. Many still lack the required funding, technology, and staff to take advantage of GIS technology. In some cases, they have banded together through coalitions, intermediaries, neighborhood GIS centers, university-community partnerships, and other institutional arrangements and relationships (Leitner et al. 1998, 2000) to resolve limitations or gaps related to technology, staff, and data limitations, and volunteer/student interns/staff turnovers. The downside to such coalitions and intermediaries is potential conflicts between neighborhood organizations about priorities (Leitner et al. 2002), and the fact that organizations are at different stages of growth—emergent versus launch/growth versus consolidation versus mature (Pathfinder International undated).

Leadership is another critical factor for the adoption and effective use of GIS and related technologies, even in the context of partnering, collaboration, and understanding how it advances the goals of the organization. Anglin and Herts (2004) note a change in community development corporation (CDC) leadership backgrounds in the post-1980 period toward more of a business model and less of an activist/social movement model.

The authors further reported that the leaders are coming to the community development field after careers in law, banking, the foundation world, and other allied fields. The implications of this changing leadership model for GIS technology adoption and sustained use is unclear, but requires further investigation beyond the scope of this paper.

Overall, the review of the literature and findings about barriers (including institutional and organizational factors) to GIS adoption by nonprofits and local government agencies reinforces the following observation made by Harris and Weiner (1998):

. . . in North America the technical problems are minor in comparison with the human ones. The success or failure of a GIS effort has rarely depended on technical factors, and almost always in institutional or managerial ones. (Harris and Weiner 1998, 68)

Against this backdrop and given concerns about the effective use of GIS by "GIS for CBOs" workshop participants, a survey was designed and administered to gain a better understanding of their community planning and land-(re)development initiatives, staffing, missions, geographic focus, GIS capacity, use of GIS, and perceived barriers to the effective use of GIS.

SURVEY METHODOLOGY

Two introduction letters (one from the Lincoln Institute's Chair of the Economic and Community Development Department and the other from the faculty researcher) were sent to 55 organizations/agencies to inform them about the research and related phone interview. Workshop #4 participants were mostly from city and development agencies, the Environmental Protection Agency (EPA), and other environmental and waste-management agencies. Given this scenario, the priority survey sample pool (n = 23) was therefore drawn primarily from the Cambridge and Tampa workshop participant list (see the Appendix for more details on workshops #1- through #4).

Staff turnover (given the workshop dates in comparison to the survey administration dates) also impacted the priority survey pool size. Table 1 was derived and confirmed using multiple indicators and sources: (l) specific information from someone else at the organization (i.e., other than the workshop participant on record), (2) bounced e-mails, and/or (3) returned postal mail.

All surveys (see Figure 1) were administered during the period of November 2006 to March 2007, and results were summarized in April and May 2007. Several participants requested that the survey be sent via e-mail, given their busy schedules in the field/community. In some instances, the organizations' mission statements were downloaded from the Internet and used as sources, in addition to responses to survey questions 3 and 4, to evaluate the limits of the mission statements with respect to community planning and development priorities, and the related use of GIS technology.

A list of ten other organizations (referred to as "other" in this paper) was compiled based on an Internet search to elicit feedback from a nationwide sample of coalitions, alliances, and intermediaries that work with or on behalf of community-based organizations, or that have similar missions and to determine the extent to which responses varied. This nationwide pool spanned several states, including Maryland (1), California (1), Minnesota (2), Michigan (1), New York (1), Indiana (1), Florida (2), and Louisiana (1).

Table 1. Staff Turnover Since Workshop

Workshop Location and Date	Percentage of workshop participants no longer with the organization on record
Cambridge, Massachusetts (March 2004)	21 %
Tampa, Florida (December 2004 and April 2005)	17 %
Atlanta, Georgia (January 2006)	8 %

Table 2. Workshop Survey Respondents' Affiliations (Self-reported)

Type of Organization	Percentage of respondents
Quasi-CDC (university-based)	6%
University Research Group	6%
Regional Development Council	6%
Community-Based Organization (CBO)	6%
Neighborhood Association	29%
Community Development Corporation (CDC)	47%

FINDINGS—WORKSHOP PARTICIPANT SURVEY SAMPLE TOOL

Survey Respondents: Type of Organization and Staffing

The majority of respondents, based on a 74 percent response rate, are affiliated with CDCs (see Table 2). Staff size, not including university-affiliated organizations[1], varies widely from zero (i.e., all volunteers) to 25. One of the organizations (with one full-time staff person) reported having three AmeriCorp members.

Organization Mission and Geographic Area of Interest

One-third of the respondents indicated community planning and land (re)development as their primary focus areas. The most popular responses (with respect to mission) were creation and preservation of affordable housing, community revitalization and restoration, and community organizing and economic develop-

1. This range does not account for the university-affiliated organizations, given that their reported numbers include the student staff.

Objective #1: Gain a better understanding of your CBO/CDC mission and geographic area of interest (as it relates to community planning and development).

1. Do you define yourself as a CBO, CDC, or neighborhood association?
2. What is your staff size?
3. What is the mission of your CBO/CDC as it relates to community planning and land development?
4. Has the mission changed in recent years?
5. How much of the day-to-day work tasks focus on community planning and land development?
6. Would you say that your geographic focus is a neighborhood, main street, region, or all of the above?
7. Who are your main collaborators for community-planning projects?
8. Do you wish to share any community-planning challenges (or successes) that stand out in your mind?

Objective #2: Gain a better understanding of GIS capacity, GIS applications for community planning, and use of data repositories.

9. Does your CBO have digital mapping/GIS capacity?
   If yes, since when? And what was the reason behind setting up GIS?
   If no, are there any plans to develop capacity in-house or otherwise?
10. How many staff members work with GIS?
11. What types of maps/data do you use most often?
12. Are you aware of useful data sources
    a. From local and regional agencies
    b. From the Internet
    c. From universities
13. Which of these datasets best enables community-level planning?
14. What about local knowledge—how is it factored into your analyses?
15. On a scale of 1-10 (with 10 being most effective), how well does your organization document project tasks/steps/analyses?

Objective #3: Understand barriers to using online data repositories and Web-GIS applications.

16. What are the pros and cons of using available online datasets/data repositories?
17. How is GIS useful (or potentially useful) to your work?
18. How is Internet-GIS useful (or potentially useful) to your work?
19. On a scale of 1-10 (with 10 being "most effective"), how effectively do you use GIS for your community-planning work?
20. What are the institutional barriers to its effective use?
21. If your organization had all the hardware, software, and data at your disposal, do you think that you could use GIS effectively for community and land-development planning?
22. Is there any other information or concerns that you wish to share?

Figure 1. Survey questions

ment. Successes seemed to revolve primarily around affordable housing assistance to low- and moderate-income residents. Others included increased participation in the planning process, with community-based organizations as lobbyists for public involvement. Specific mission statements gleaned from Web sites and other documents submitted by respondents offered broader agendas such as enhancement of quality of life, fostering of civic engagement and community leadership, provision of economic opportunities, neighborhood planning, and historic preservation.

Diminished funding sources and difficulty with keeping neighborhood projects as a priority for multiyear funding cycles and the general lack of funding for smaller communities were two community-planning challenges raised. The planning process also featured in the list of concerns, but mainly with respect to getting parties to the tables and keeping them involved and interested, as well as convincing the city administration to provide technical guidance and direction to ensure that neighborhood plans are practical and implementable. Related concerns about the difficulty with accessing information on real-time availability of land and bridging the information gap between the university and community were also raised.

The organizations' geographic focus also varies, capturing neighborhood, city, and regional boundaries (see Table 3), and multiple collaborators were acknowledged (including neighborhood groups and residents, city/local government, planning departments, other CBOs and CDCs, state agencies, universities, nonprofit agencies, and developers of proposed projects).

GIS Capacity, GIS Applications and Effective Use

Fifty-three percent of the respondents reported having mapping/GIS capacity, with half of these respondents giving their organization a rating of 8 or higher (on a scale of 1-10, with 10 being "most effective") in response to the question about the effectiveness level of using GIS for community-planning work. There were no specific characteristics of this subgroup, although its use of older ESRI software (ArcView 3.2) and Microsoft's MapPoint further highlight the fact that the newer popular GIS software is cost-prohibitive for many community-based organizations.

The main reasons for initial GIS adoption were self-reported as organizing data, geocoding/address matching, generating maps and reports, and performing analysis with respect to potential development sites and comprehensive planning. Respondents suggested a much broader list of potential use, however, including

Table 3. Workshop Survey Respondents' Geographic Focus (Self-reported)

Geographic Area	Percentage of respondents
Main Street	0%
City	6%
Region	18%
Neighborhood	64%
All of the above	12%

more effective communication and documentation of community assets and liabilities, visualization (including current and projected neighborhood and regional land-use and sociodemographic patterns), site identification and feasibility analysis, tracking property disposition as part of a community-based housing code enforcement system, and enhancing grant and funding proposals.

Respondents reported commonly used data for performing these analyses, including census data, city and neighborhood boundaries, parcels, land use, zoning, roads, housing statistics, membership data, tax-assessment data, vacant-housing inventories, business patterns, crime statistics, housing sales, and digital imagery. They appreciated the free and accessible online data and GIS resources, but these resources were generally deemed inadequate for the following reasons: unreliability, obsolescence, inaccuracy, technical knowledge requirement, lack of context, the inability to easily manipulate and tailor to their needs, and use of different categories by users/neighborhood representatives, which, in turn, can lead to faulty comparisons. According to one interviewee, organizations will increasingly use more progressive and appropriately designed tools such as Google Maps and Google Earth. Another interviewee looked forward to the improved functionality of open-source software as a viable and more affordable option for community-based organizations.

Institutional and Organizational Barriers

In conducting the survey, the interest in identifying institutional and organizational factors and barriers that impact the effective use of GIS and related technologies for community planning and land (re)development was emphasized. Reported constraints and barriers varied widely but can be grouped into four main categories (staffing, utility and relevance, mission mismatch, and other general resource constraints), which are outlined below.

Staffing was listed as a common barrier (i.e., the size of the organization does not allow for having a dedicated GIS person, especially in relation to other time-sensitive and mission-critical tasks and the constant juggling of budgets, projects, and resources). The reliance on a large volunteer pool, which can mean that GIS projects are undertaken by individuals passionate about specific projects, was listed as a related barrier. Even with a GIS champion (whether full-time staff person or volunteer), the difficulty with demonstrating and convincing other staff members about the potential utility and relevance was a commonly cited problem. In instances when training was encouraged, it was mostly on a one-time basis, with lack of dedicated funds for additional training and technical assistance.

Lack of a purpose or mission that promotes use of GIS for community planning was raised as a barrier, in addition to a lack of appropriate connections, collaborators, and networking to organizations that can provide data or answer GIS-related questions. Conflicting views were also presented about community-university partnerships, where a lack of sustained support from university research group partners was reported by a CBO workshop participant, while the lack of stable leadership of CBOs was offered as a barrier by a respondent from a university.

As expected, hardware, software, and training costs were still presented as barriers, but with an interesting caveat by one respondent that infrastructure costs became an additional cost for CBOs with temporary office locations that generally lack needed infrastructure.

FINDINGS—"OTHERS" SURVEY SAMPLE POOL

Survey Respondents: Type of Organization and Staffing

The majority of respondents, based on a 60 percent response rate, are affiliated with CBOs (see Table 4), with a staff size that ranges from 5 to 15.

Organization Mission and Geographic Area of Interest

The responses (with respect to mission) varied widely—provision of indicators to neighborhood associations; assistance with technology needs and program evaluation; revitalization of areas in need; promotion and stimulation of residential, commercial, and economic revitalization; and empowering residents to collaborate with other entities to develop neighborhood action plans. The main challenge reported is decision making about which projects to take up that ensure equity among neighborhoods.

Community planning and land (re)development is the primary day-to-day focus for 50 percent of the organizations interviewed, and for 83 percent of the organizations, the geographic focus captures multiple neighborhood, city, and regional boundaries. Collaborators include neighborhood groups and associations, nonprofits and foundations, local government, private companies, and other community-based organizations.

GIS Capacity, GIS Applications, and Effective Use

Sixty-six percent of respondents reported having mapping/GIS capacity, and staff members with working knowledge of GIS ranged from 17 percent to 25 percent. The main reasons for initial GIS adoption were self-reported as organizing data (including census data, city and neighborhood boundaries, parcels and housing statistics), geocoding/address matching, or as evolving from a

Table 4. "Other" Survey Respondents' Affiliations (Self-reported)

Type of Organization	Percent of respondents
Quasi-CDC (university-based)	0%
University Research Group	0%
Regional Development Council	0%
Community-Based Organization (CBO)	50%
Neighborhood Association	16.67%
Community Development Corporation (CDC)	0%
Resource Center	16.67%
Community Redevelopment Corporation (CRC)	16.67%

university partnership. However, they listed visualization, communication, and analysis as potential uses of GIS for community planning and land (re)development.

Of the respondents with GIS capacity, 75 percent gave their organizations a rating of 8 or above (i.e., on a scale of 1-10, with 10 being "most effective") in response to the question about the effectiveness level of using GIS for community-planning work.

Institutional and Organizational Barriers

Reported constraints and barriers relate to a lack of leadership embracement and a general lack of understanding about GIS functionality and its applicability to existing projects, community planning, development, and other day-to-day activities. This is compounded by concerns about how to use and analyze the data in an ethical manner, including the preservation of confidentiality when using "sensitive" data. An understated but important barrier reported was the ability of organizations to survive without GIS.

By design, the "Others" survey sample pool focuses on broader regional and coalition initiatives such as revitalization of areas in need, provision of indicators, and assistance with technology needs compared to the focus of workshop participants and their representative organizations (i.e., CDCs, CBOs, and neighborhood associations focused on local issues such as affordable housing, community revitalization and restoration, community organizing, and economic development).

The key differences in responses relate to geographic scope and self-report effectiveness of using GIS for community planning and development (see Table 5).

Table 5. Workshop Participant Versus Others: Comparing Sample Responses

	Participants	Others
GIS capacity	53%	66%
Staff members with working knowledge of GIS	0–50%	17–25%
Effectiveness of using GIS (8 or greater)	24%	75%
Geographic focus	64% (neighborhood)	83% (neighborhood, city, and region)

CONCLUSIONS AND IMPLICATIONS

All survey respondents strongly or somewhat agreed that given the right hardware, software, and data at their disposal, they could use GIS effectively for community and land-development planning. However, there was a clear mismatch between the actual use and the potential use of GIS. That is, despite the actual use and application of GIS technology being limited, the potential uses were extremely well articulated, far outweighing and outnumbering the uses reported by respondents. The focus of community-based organizations on a neighborhood/community unit of analysis, while typical, is not insignificant given issues of data obsolescence, and the difficulty of easily incorporating local knowledge. CBOs, with the least amount of resources and GIS expertise, are essentially faced with conceptual, contextual, and complex problems difficult even for people conducting such neighborhood-level analyses for decades. A look at the long-term multicomponent and multi-investigator nature of public participation GIS initiatives is partial proof of this unresolved complexity. Staffing in community-based organizations presents an additional set of challenges. At the time of the surveys, for example, approximately one-fifth of the participants from the Cambridge and Tampa "GIS for CBOs" workshops organized by the Lincoln Institute of Land Policy were no longer with the organization on record.

Although a long list of barriers (resource, organizational, and institutional) are reported in earlier sections of this paper, three of them stand out:

1. Difficulty in applying technology to existing projects,
2. Lack a mission statement that promotes the use of IT (in general) for community planning, and
3. Ability of CBOs to survive without GIS.

All observations from these research findings, workshop development, and prior experiences with community-based organizations point to the need for a more realistic organizational and management framework promoted by Huxhold and Levinsohn (1995), rather than the widely used "hardware, software, data, and people" framework. For community-based organizations, three foci are recommended: organizational capacity, context, and analysis:

Organizational capacity. Clarify (1) the connection between organizations' missions and technology use/application for benefits to community planning and redevelopment processes and (2) how GIS fits into the existing day-to-day operations of a CBO assuming typical staff expertise. Future research on CBO leadership models (i.e., activist versus business model) to assess whether any one model is a predictor of embracement and the sustained and effective use of GIS and related technologies is also beneficial in this regard.

Context. Outline the GIS analytical processes and products up front. CBOs should initially identify one or two specific land-policy/land-development questions that can be used as context for organizing GIS data resources and for crafting comprehensive GIS-related project and workflow.

Analysis. Selection, manipulation, and analysis of data from a variety of sources using spatial-analysis techniques and map-making to communicate results remains at the core of any GIS project. Issues related to ethics, confidentiality, and sensitive data must also be addressed.

This proposed framework can also be used to reframe the content of customized professional development workshops and training. It is important to have a diverse faculty group, but with CBO representatives playing a key role in instruction. In that regard, CBO faculty who focus 100 percent of their efforts on community planning and redevelopment, have similarly low GIS staff ratios, and face typical barriers voiced by survey respondents should be sought. On the participant end, CBO attendees can benefit from inclusion of their collaborators (i.e., public officials, city government, universities, developers, and volunteers). The dialogue and appreciation for what each other does is important.

Taken out of context, these findings can appear daunting for funders, university collaborators, and other agencies interested in promoting the work performed by community-based organizations and advocating the importance of GIS and other information technologies to enhance their work, as organizers/hosts of ongoing professional development workshops and training sessions. However, the research simply highlights a complex set of factors and barriers that continue to lead to the untapped use of GIS and related technologies by community-based organizations. The bottom line is that a careful look at organizational structure and capacity is a vital aspect when introducing or applying new technologies. It is also important to acknowledge that GIS is not always necessary for CBO day-to-day activities. This can help avoid the common problem of temporary GIS champions creating more disruption, an unintentional setback that Sieber (2000) warned against several years ago.

Appendix: Workshop Location, Focus, and Participants

Following is a brief overview of each workshop, taught by a group with a wide range of expertise and affiliations (i.e., university, CBO, research and advocacy, and private-sector GIS vendor). In addition to brochures and course announcements posted on the Lincoln Institute's Web site, workshop announcements were mailed and e-mailed to CBOs. Workshops were offered for a minimum cost (less than $50 in most cases to gauge commitment to attending), but scholarships/fee waivers were provided. In Workshops #2, #3, and #4, multiple attendees represented the same organization.

Workshop #1 (Cambridge, Massachusetts, March 2004): Nineteen participants registered for the first workshop held in Cambridge, Massachusetts, at the Lincoln Institute's headquarters. The workshop drew participants from Massachusetts, Maine, New York, New Hampshire, Pennsylvania, Florida, and Oregon. The course was held over a period of two days, with one day dedicated to a site visit and tour of the Dudley Street Neighborhood Initiative (DSNI) main office.

Workshops #2 and #3 (Tampa, Florida, December 2004 and April 2005): The December 2004 workshop was the first workshop focused on a specific region. A total of 42 participants registered for the two workshops in Tampa, Florida. The majority of participants were from CBOs or neighborhood associations in the Tampa Bay region and south Florida. Several had recently taken a civic leadership course organized by the Jim Walter Partnership Center (University of South Florida), also a cosponsor of the Tampa workshops.

The format for the April 2005 workshop varied slightly. There were two concurrent morning sessions—one offering hands-on training in ArcGIS software (based on the popular request from December 2004 participants and sponsored by Environmental Systems Research Institute) and the other covering December 2004 material for new participants. The afternoon session brought together participants from both morning sessions to interact and collectively explore GIS implementation and capacity-building issues relevant to the Tampa Bay region.

Workshop #4 (Atlanta, Georgia, January, 2006): Thirty-four participants registered for the one-day "GIS for CBOs" workshop, which followed another one-day workshop on the topic of "Recycling Vacant, Abandoned, and Contaminated Properties," also organized by the Lincoln Institute of Land Policy.

Acknowledgments

The feedback and survey responses of representatives from community-based organizations formed the basis of the reported findings. Their time, patience, and contributions were invaluable and are very much appreciated. The comments and feedback provided by the reviewers were extremely useful in revising the paper. Macline Pierre worked as a graduate research assistant on this project and was instrumental in collecting and compiling the survey responses. Most important, this assessment would not have been possible without funding from the Lincoln Institute of Land Policy and the leadership provided by Rosalind Greenstein, Chair of the Institute's Department of Economic and Community Development.

About the Author

Dr. Ann-Margaret Esnard is a faculty member in the Department of Urban and Regional Planning and Director of the Visual Planning Technology Lab at Florida Atlantic University, Fort Lauderdale. Esnard is a faculty associate at Lincoln Institute of Land Policy and was the course developer for the Institute's "GIS for CBOs" workshop series (2004–2006).

Corresponding Address:
Department of Urban and Regional Planning
Florida Atlantic University
Fort Lauderdale, FL 33301
954-762-5646
Fax: 954-762-5683
aesnard@fau.edu

References

Al-Kodmany, K. 2000. Extending geographic information systems to meet neighborhood planning needs: the case of three Chicago communities. URISA Journal 12: 19-37.

Anglin, R. V., and R. D. Herts. 2004. Limitations to organizational and leadership progress in community development: an overview. In Anglin, R. V., Building the organizations that build communities: strengthening the capacity of faith- and community-cased development organizations. U.S. Department of Housing and Urban Development: Office of Policy Development and Research, 13-26.

Boyd, S., and R. Chan. 2002. Placemaking: tools for community action. Accessed May 2007, http://www.sustainable.org/Placemaking_v1.pdf.

Brumley, R. D. 2003. TriCentral PC toolkit: Chapter 3, Building a palliative care program—identify barriers to implementation and strategies to overcome. Accessed May 2007, http://www.growthhouse.org/palliative/toolkit.html.

Chandler, A., G. Thomas Kingsley, J. Kirschenbaum, and K. L. S. Pettit. 2006. The potential of parcel-based GIS in community development and urban land management. Cambridge, MA: Lincoln Institute of Land Policy Working Paper.

Craig, W., T. M. Harris, and D. Weiner, Eds., Community participation and Geographic Information Systems. London: Taylor and Francis, 232-45.

Elwood, S., and H. Leitner. 1998a. GIS and community-based planning: exploring the diversity of neighborhood perspective and needs. Cartography and Geographic Information Systems 25: 77-88.

Elwood, S., and H. Leitner. 1998b. GIS and spatial knowledge production for neighborhood revitalization negotiating state priorities and neighborhood vision. Journal of Urban Affairs 25(2): 139-57.

Esnard, A.-M., M. Gelobter, and X. Morales. 2004. Environmental justice, GIS and pedagogy. Cartographica 38(3&4): 53-61.

Glickman, N. J. 2004. Building capacity of community-based development organizations: the case of community development partnerships. In Anglin, R. V., Building the organizations that build communities: strengthening the capacity of faith- and community-based development organizations. U.S. Department of Housing and Urban Development: Office of Policy Development and Research, 117-26.

Glickman, N. J., D. Devance-Manzini, and S. DiGiovanna. 2004. Expanding organizational capacity: the human capital development initiative. In Anglin, R. V., Building the organizations that build communities: strengthening the capacity of faith- and community-based development organizations. U.S. Department of Housing and Urban Development: Office of Policy Development and Research, 165-74.

Greenstein, R. 2006. How can community assets be best mobilized to benefit community? Lincoln Institute of Land Policy workshop brochure, August 26 and 27, 2006, Atlanta, Georgia.

Harris, T., and D. Weiner. 1998. Empowerment, marginalization, and "community-integrated" GIS. Cartography and Geographic Information Systems 25(2): 67-76.

Huxhold, W. E., and A. G. Levinsohn. 1995. Managing Geographic Information System projects. New York: Oxford University Press.

Innes, J. E., and D. M. Simpson. 1993. Implementing GIS for planning: lessons from the history of technological innovation. Journal of the American Planning Association 59(2): 230-36.

Kwaku Kyem, P. A. 2002. Promoting local community participation in forest management through application of a Geographic Information System: A PPGIS experience from southern Ghana. In Craig, W., T. M. Harris, and D. Weiner, Eds., Community participation and Geographic Information Systems. London: Taylor and Francis, 232-45.

Leitner, H., R. B. McMaster, S. Elwood, S. McMaster, and E. Sheppard. 2002. Models for making GIS available to community organizations: dimension of difference and appropriateness. In Craig, W., T. M. Harris, and D. Weiner, Eds., Community participation and Geographic Information Systems. London: Taylor and Francis, 37-52.

Leitner, H., S. Elwood, E. Sheppard, S. McMaster, and R. McMaster. 2000. Modes of GIS provision and their appropriateness for neighborhood organizations—examples from Minneapolis and St. Paul, Minnesota. URISA Journal 12(4): 43-56.

LISC and PolicyLink. 2002. Mapping for change: using GIS for community development. Accessed May 2007, http://www.lisc.org/content/publications/detail/835.

Mayer, N. S. 2004. Education and training for community development. In Anglin, R. V., Building the organizations that build communities: strengthening the capacity of faith- and community-based development organizations. U.S. Department of Housing and Urban Development: Office of Policy Development and Research, 249-69.

Nedovic-Budic, Z., and J. K. Pinto. 1999. Understanding interorganizational GIS activities: a conceptual framework. URISA Journal 11(1): 53-64.

Pathfinder International. Undated. Organizational structure: organizational development module 2. Accessed May 2007, http://www.pathfind.org/siteDocServer/.Organizational_Structure.complete.pdf?docID=323.

Ramasubramanian, L. 1999. GIS implementation in developing countries: learning from organisational theory and reflective practice. Transactions in GIS 3(4): 359-80.

Ramos, R. 2001. Community incentives and capacity building for CBPR: successfully promoting community interests through research. Presented at a conference on Community-Based Participatory Research, November 27-28, 2001. Accessed May 2007, http://www.ahrq.gov/about/cpcr/cbpr/cbpr2.htm.

Rattray, N. 2006. A user-centered model for community-based Web-GIS. URISA Journal 18(2): 25-34.

Rodriguez, A., and N. Herzog. 2004. Replacing passionate leaders: the current challenge for community development. In Anglin, R. V., Building the organizations that build communities: strengthening the capacity of faith- and community-based development organizations. U.S. Department of Housing and Urban Development: Office of Policy Development and Research, 93-116.
Roman, C. G., and G. E. Moore. 2004. Measuring local institutions and organizations: the role of community institutional capacity in social capital. Washington, D.C.: The Urban Institute Justice Center.
Sieber, R. E. 2000. GIS implementation in the grassroots. URISA Journal 12(1): 15-29.
Somers, R. 1998. Developing GIS management strategies for an organization. Journal of Housing Research 9 (1): 157-77.
Stonich, S. 1998. Information technologies, advocacy, and development: resistance and backlash to industrial shrimp farming. Cartography and Geographic Information Systems 25(2): 113-22.
Talen, E. 2000. Bottom-up GIS: a new tool for individual and group expression in participatory planning. Journal of the American Planning Association 66(3): 279-94.
Ventura, S. 1995. The use of geographic systems in local government. Public Administration Review 55(5): 461-67.

Rely on Rolta to

Activate

Your Geospatial Assets

CELEBRATING 25 Years

High-End GIS and GeoSpatial Services

- Conversion and Migration Services
- Web Enabled GIS Solutions
- Digital Photogrammetry and Image Processing
- GIS / Spatial Data Management
- Mapping and Enterprise Application Integration for
  - Environmental Information Systems
  - Public Safety & Emergency Services
  - Land Record / Natural Resources Management
  - Urban Planning / Land Use Studies
  - Economic Development
  - Election Management and Mapping
  - Terrain Visualization
- Spatial Application management and consulting

ISO 9001 : 2000 CERTIFIED

OnPoint and OnSite For Community Development

OnPoint™ is used by local governments across North America in promoting economic development throughout their communities. Our **OnSite™** software (an extension to **OnPoint™**) is specifically designed to help businesses interested in your community find the perfect location for their business.

OnSite™ allows members of the business community to browse potential properties and then drill down on particulars about the site, the area's demographics, the infrastructure, and the zoning. Users can print reports that summarize the data they discover. The entire process can be controlled completely through the **OnSite™** administration tool, which allows the GIS specialists to customize the search options and contextual information presented to the end user to maximize efficiency and leverage the unique characteristics of the area.

ESRI ORACLE Microsoft GOLD CERTIFIED Partner ORION A ROLTA COMPANY

Rolta International, Inc.
5865 North Point Pkwy.
Alpharetta, GA 30022
(678) 942-5000

Email: Sales@RoltaUS.com

https://www.roltaus.com/URISA

ROLTA

Coupling Multiagent Geosimulation and Spatial OLAP for Better Geosimulation Data Analysis

Walid Ali, Bernard Moulin, Yvan Bédard, Marie-Josée Proulx, and Sonia Rivest

Abstract: *Building a good simulation model can be a great deal of work. We need to specify and model the system to be simulated, implement the model, collect data on the corresponding real system (if any), verify and validate the simulation system, and run the simulation. To test different ideas, to learn about the system's behavior in new situations, and to make efficient decisions, decision makers need ways to explore the simulation outputs easily and rapidly. Geosimulation users need tools to analyze the spatial output data. In this paper we present an approach that combines spatial online analytical processing (SOLAP) techniques with multiagent geosimulation techniques to improve the exploration of spatial and nonspatial data resulting from geosimulations. As an example, we present the application of our approach to the simulation of the shopping behavior of customers in a shopping mall.*

INTRODUCTION AND MOTIVATIONS

Simulation output analysis is a very relevant step in a simulation study. This step is necessary to test different ideas and to learn about the simulation model and the corresponding simulation system. A user needs to better understand the simulation model's output and, consequently, needs appropriate techniques to analyze the simulation output. Some researchers (Seila 1992, Alexopoulos et al. 1998, Alexopolous 2000, Sanchez 2001) propose different analysis techniques for simulation outputs. Decision makers usually find the statistical and mathematical aspects of these analysis techniques difficult to use. To overcome this difficulty, Grier (1992) proposed a graphical statistical analysis technique that can be used to analyze simulation outputs. The visual display of the results quickly conveys information about the simulation model. Users who rely on simulations to support their decisions prefer graphical analyses because they are easy to understand. Blaisdell (Blaisdell et al. 1992) proposed SIMSTAT, a tool used to analyze simulations based on a statistical graphical analysis technique that is combined with several discrete-event simulation tools. Using statistical graphical analysis is efficient for event-discrete simulations or other kinds of simulations that do not deal with spatial or geographic data. However, for certain geosimulation (GS) fields such as urban simulation (US) or multiagent geosimulation (MAGS), spatial data become an important issue for decision makers. In such simulation fields, a classical analysis technique based on tables and statistical graphs is too limited for spatial analysis: no spatial analysis (*e.g.,* correlation between spatial variables), no spatial visualization, no map-based or cartographical exploration of spatial data (which is relevant to reveal clusters, the proximity between two phenomena or the spatial area of a phenomenon), etc. Multiagent geosimulation (MAGS) has a great potential when explaining the subtle interactions of heterogeneous actors in complex social systems taking into account the geographic aspect of the simulation environment. The characteristics of the agents (autonomy, social ability, proactiveness, advanced spatial behavior such as perception, navigation, and memorization, etc.) and the spatial features of the simulation environment make MAGS an attractive approach to develop simulations of complex systems. In addition, the complexity of the simulation models and their visualization capabilities (cartographical visualization, 2-D–3-D displays) make them more realistic and, therefore, closer to users' mental models. Generally, simulation applications generate outputs, which need to be interpreted by the users. The huge volume and the complexity of these outputs, a big part of which is spatial, make them difficult to be interpreted and analyzed by the users. To analyze these outputs of simulations that are performed in spatial environments, we need more sophisticated analysis techniques, which can be used to analyze and explore complex simulation models and outputs involving geographic data. These analysis techniques must also be compatible with users' mental models and must generate analysis results that can be easily exploited by users. Consequently, the traditional statistical or mathematical analysis techniques are less suited to analyze geosimulation outputs because they are less efficient to analyze and explore the spatial aspects of the simulation outputs, which are very important for a geosimulation. After an in-depth comparison of several analysis techniques and tools (Seila 1992, Blaisdell et al. 1992, Grier 1992, Kelton 1997, Alexopolous et al. 1998, Sanchez 2001, Alexopolous 2002), we found that the most appropriate one for simulation outputs is the online analytical process (OLAP) because: (1) OLAP allows users and analysts to explore data in the way they think, across multiple variables, called *dimensions,* at the same time (Codd et al. 1993); (2) the multidimensional approach of OLAP analysis is more in agreement with the end user's mental model (Yougworth 1995); and (3) we can take into account the spatial aspect of the data to be analyzed using a recent extension of OLAP called spatial online analytical processing (SOLAP) (Bédard 1997).

In this paper we present the fundamental concepts of the geosimulation field and we discuss how we can conduct nonspatial and spatial analysis of a multiagent geosimulation output using the SOLAP multidimensional analysis techniques. Based on a geosimulation prototype, we show how end users can easily and rapidly use these techniques to analyze and explore both nonspatial and spatial outputs of geosimulations.

In the following two sections, we present fundamental notions related to the geosimulation and OLAP–SOLAP analysis techniques. The next section shows how decision makers can use OLAP–SOLAP techniques to make better nonspatial and spatial analysis of a geosimulation output. The final two sections discuss related works and present conclusions.

GEOSIMULATION, MULTIAGENT GEOSIMULATION, AND THEIR APPLICATIONS

This section aims at presenting the basic concepts characterizing the geosimulation and multiagent geosimulation (MAGS) fields.

Geosimulation and Multiagent Geosimulation

Geosimulation is concerned with the design and the construction of high-resolution spatial models, using these models to explore ideas and hypotheses about how spatial systems operate, developing simulation software and tools to support object-based simulation, and applying simulation to solve real problems in geographic contexts (Benenson and Torrens 2004). Geosimulation differs from conventional urban simulation in its constituent "elements." Geosimulation models operate with human individuals and infrastructure entities, represented at spatially nonmodifiable scales such as households, homes, or vehicles. Many of these entities are animated (visually and dynamically), and that animation drives the behavior of inanimate objects in a simulation (Benenson and Torrens 2004). Geosimulation is a useful tool to integrate the spatial dimension in models of interactions of different types (economical, political, social, etc.) (Mandl 2000).

Mand, Koch, and Moulin (Mand 2000, Koch 2001, Moulin et al. 2003) presented multiagent geosimulation as a coupling of two technologies: the multiagent-based simulation (MABS) technology and the geographic information system (GIS). Based on the MABS technology, the simulated entities are represented by software agents that can be autonomous in their behavior. They can interact with other agents and with the spatial environment. They may be active, reactive, mobile, social, or cognitive (Koch 2001). Thanks to the agents' capabilities, we can use them to model and simulate complex entities or systems.

Using GIS technology, spatial features of geographic data can be introduced in the simulation. The GIS plays an important role in the development of geosimulation models. New methodologies for manipulating and interpreting spatial data developed by geographic information science and implemented in GIS have created added-value for these data (Benenson et al. 2003).

Multiagent geosimulation is a powerful concept that can be used to simulate complex systems in georeferenced environments. According to our literature review, a small number of multiagent geosimulation platforms can be used to simulate systems in geographic environments using the agent paradigm. For example, we can cite the following platforms: common-pool resources and multiagent systems (CORMAS) (Bousquet et al. 1998), multiagent geosimulation (MAGS) (Moulin et al. 2003), and RePast (Dibble and Feldman, 2004) (Najlis and North, 2004). In our work, we use the MAGS platform to simulate the shopping behavior in a mall.

THE MAGS PLATFORM

MAGS is a generic platform that can be used to simulate, in real time, thousands of knowledge-based agents navigating in a two-dimensional (2-D) or three-dimensional (3-D) virtual environment. MAGS agents have several knowledge-based capabilities such as perception, navigation, memorization, communication, and objective-based behavior that allow them to display an autonomous behavior within a 2-D–3-D geographic virtual environment (Moulin et al. 2003). The agents in MAGS are able to perceive the elements contained in the environment, to navigate autonomously inside it, and to react to changes occurring in the environment. These agents have several knowledge-based capabilities (Moulin et al. 2003) that are briefly presented in the following paragraphs.

The agent perception process. In MAGS, agents can perceive (1) terrain characteristics such as elevations and slopes; (2) elements contained in the landscape surrounding them, including buildings and static objects; (3) other mobile agents navigating in the agent's range of perception; (4) dynamic areas or volumes whose shape changes during the simulation (for example, smoky areas or zones having pleasant odors); (5) spatial events such as explosions, etc., occurring in the agent's vicinity; and (6) messages communicated by other agents (Moulin et al. 2003).

The agent navigation process. In MAGS, agents can have two navigation modes: *following-a-path mode* in which agents follow specific paths, which are stored in a bitmap called ARIADNE_MAP, or *obstacle-avoidance mode* in which the agents move through open spaces avoiding obstacles. In MAGS, the obstacles to be avoided are recorded in a specific bitmap called OBSTACLE_MAP.

The memorization process. In MAGS, the agents have three kinds of memory: *perception memory* in which an agent stores what it perceived during the past few simulation steps; *working memory* in which the perceived data is filtered according to the agent's interests (or focus of attention); and *long-term memory* in which the agent stores what is worth recording after the simulation completion (Perron et al. 2004).

The agent's characteristics. In MAGS, an agent is characterized by a number of variables whose values describe the agent's state at any given time. *Static states* and *dynamic states* are distinguished. A static state does not change during the simulation and is represented by a variable and its current value (for instance, gender, age group, occupation, marital status). A dynamic state is a state that can possibly change during the simulation (for example, hunger, tiredness, stress). A dynamic state is represented by a variable associated with a function that computes how this variable changes values during the simulation. The variable is characterized by an initial value, a maximum value, an increase rate, a decrease rate, an upper threshold, and a lower threshold that are used by the function. Using these parameters, the system can simulate the evolution of the agent's dynamic states such as the agent's needs and trigger the relevant behavior (Moulin et al. 2003).

The objective-based behavior. In MAGS, an agent is associated with a set of objectives that it tries to reach. The objectives are organized in hierarchies, which are trees composed of nodes representing composite objectives and leaves representing elementary objectives that are associated with actions that the agent can perform. Each agent owns a set of objectives corresponding to its needs. An objective is associated with rules containing constraints on the activation and the completion of the objective. Constraints depend on time, on the agent's states, and the environment's state. The selection of the current agent's behavior relies on the priority of its objectives. Each need is associated with a priority, which varies according to the agent's profile. An objective's priority is primarily a function of the corresponding need's priority. It is also subject to modifications brought by the opportunities that the agent perceives or by temporal constraints (Moulin et al. 2003).

The agent communication process. In MAGS, agents can communicate with other agents by exchanging messages using mailbox-based communication.

In MAGS, the environment is very important; the spatial characteristics of the environment and static objects are generated from data stored in a geographic information system and in related databases. The spatial characteristics of the environment are recorded in raster mode, which enables agents to access the information contained in various bitmaps that encode different kinds of information about the virtual environment and the objects contained in it. The *AgentsMap* contains the information about the locations of agents and of the static objects contained in the environment. The *ObstaclesMap* contains the locations of obstacles, the *AriadneMap* contains the paths that can be followed by mobile agents, the *HeightMap* represents the elevations of the environment, etc. The information contained in the different bitmaps influences the agent's perception and navigation. In MAGS, the simulation environment is not static and can change during the simulation. For example, we can add new obstacles or gaseous phenomena such as smoke, dense gases, and odors, which are represented using particle systems, etc. (Moulin et al. 2003).

As examples of multiagent geosimulation applications, we can cite a project that simulates the pedestrian behavior in Quebec City (Moulin et al. 2003) and an application that simulates the shopping behavior in Square One mall in Toronto (Ali and Moulin 2005). These two applications use the MAGS system as a simulation platform (Moulin et al., 2003). In the following section, we present the shopping mall simulation application (Ali and Moulin 2005).

Mall_MAGS: Multiagent Geosimulation of Shopping Behavior in a Mall

The Mall_MAGS prototype is a multiagent geosimulator that simulates human shopping behavior in a shopping mall. This simulator is developed using the MAGS platform (Moulin et al. 2003). Human shopping behavior is influenced by several parameters that come from the shopper himself or herself (internal factors) and from the environment (external factors). Among the external factors are several spatial factors: the layout of the stores or kiosks, the spatial configuration of the shopping mall, the store locations, etc. To better understand customers' shopping behavior in a shopping mall, we try to simulate it. To understand the influence of one simulation variable on another, we need to apply adequate data analysis on the input/output simulation data.

Using the MAGS platform, we developed a multiagent geosimulation prototype that simulates customers' shopping behavior in a virtual mall. As a case study, we used the Square One shopping mall in Toronto, Canada. To feed the simulation models with real data, we carried out a survey in October of 2003 and collected 390 questionnaires filled out by real shoppers in the Square One shopping mall. This data belongs in two categories: nonspatial data such as demographic information (gender, age group, marital status, occupation, preferences, habits, etc.) and spatial data such as preferred entrance and exit doors, habitual itineraries, well-known areas in the mall, etc. We also used spatial data for the simulation environment; this data comes from a geographic information system representing the Square One mall (see Figure 1).

In Figures 2a and 2b, we present 2-D and 3-D screenshots of a simulation that involved 390 Shopper software agents navigating in the virtual shopping mall in 2-D and 3-D. In the 2-D simulation, each store is represented by a color in red-scale colors, which have no significance and are used only to distinguish

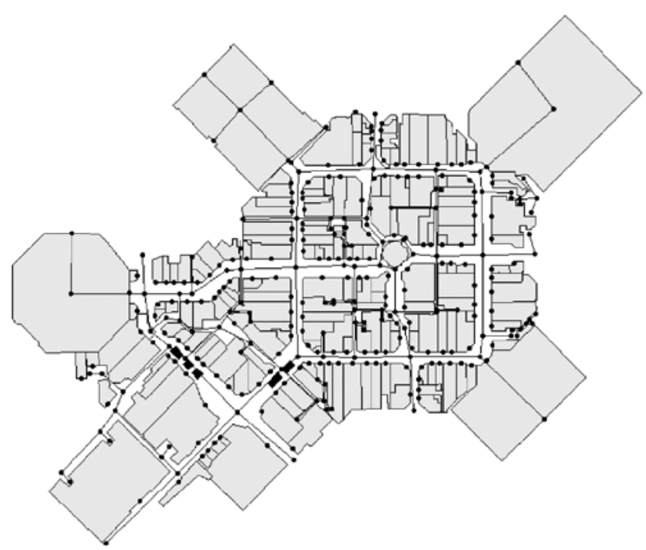


Figure 1. The 2-D spatial structure of Square One shopping mall where points indicate store entrances and polygons illustrate stores

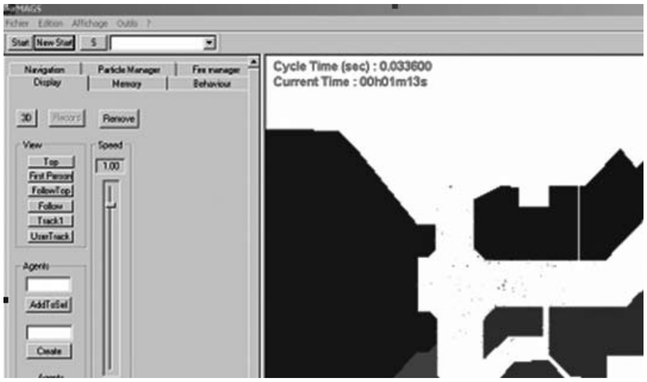


Figure 2a. The 2-D simulation in MAGS platform (Square One mall)

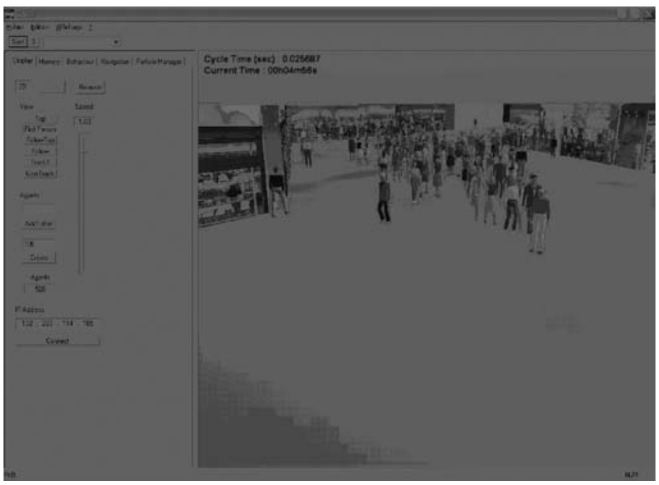


Figure 2b. The 3-D simulation in MAGS platform (Square One mall)

between the stores.

In the simulation prototype, the Shopper agent comes to the mall to visit a list of specific stores or kiosks that are chosen before the simulation on the basis of the agent's characteristics and according to the survey data. It enters by a particular door and starts its shopping trip. Based on its position in the mall, its knowledge (memorization process), and what it perceived in the mall (perception process), it chooses the next store or kiosk to visit (decision-making process). When a store or kiosk is chosen, the agent moves in its direction (navigation process). Sometimes, while it is moving to the chosen store or kiosk, the agent perceives another store or kiosk (perception process) that is in its shopping list and whose location it did not know a priori. In this case, the Shopper agent moves to this store or kiosk and memorizes it (memorization process) for its subsequent shopping trips. The Shopper agent pursues its shopping trip until it visits all the stores or kiosks on its list or until it has no time left. If the Shopper agent still has time for shopping and some stores or kiosks on its list are in locations unknown by the agent, it starts to explore the shopping mall to search for these stores or kiosks. When the Shopper agent reaches the maximum time allowed for the shopping trip, it leaves the mall.

The Shopper agent can also come to the mall for exploration purposes, without any specific list of stores or kiosks to visit. In the exploration mode, the Shopper agent takes its preferred paths in the shopping mall. In this model, the Shopper agent's movements toward stores, kiosks, music zones, odor zones, or lighting zones are influenced by its habits and preferences. For example, if the Shopper agent likes cars and it passes in front of a car exhibition, it can attend this exhibition. To extend our simulation prototype, we could simulate the Shopper agent's reactions to the mall's atmosphere. We can insert special ambience agents that broadcast music, lighting, or odor. If the Shopper agent is in the exploration mode and likes the music, the lighting, or the odor broadcasted by these ambience agents, the Shopper agent can move toward them and possibly enter the related store.

During its shopping trip, the Shopper agent can feel the need to eat or to go to the restroom (simulated by a dynamic variable reaching a given threshold). Because these needs have a bigger priority than the need to shop, the agent temporarily suspends its shopping trip and goes to the locations where it can eat something or go to a restroom. In our geosimulation prototype, the priorities of the activities of the shopping behavior are defined based on Maslow's hierarchy of needs (Maslow 1970).

THE OLAP AND SOLAP DATA ANALYSIS TECHNIQUES

This section presents the fundamental concepts on which the OLAP and SOLAP analysis techniques are based. These techniques are used to analyze the simulation output data generated by the geosimulation prototype that simulates the shopping behavior in a mall.

OLAP Analysis Technique

Online Analytical Processing (OLAP) has been defined as ". . . the name given to the dynamic enterprise analysis required to create, manipulate, animate and synthesize information from exegetical, contemplative and formulaic data analysis models. This includes the ability to discern new or unanticipated relationships between variables, the ability to identify the parameters necessary to handle large amounts of data, to create an unlimited number of dimensions, and to specify cross-dimensional conditions and expressions" (Codd et al. 1993). Other OLAP definitions have since been proposed, including "A software category intended for the rapid exploration and analysis of data based on multidimensional approach with several aggregation levels" (Caron 1998).

The multidimensional approach is based on *dimensions* and *measures*. *Dimensions* represent the analysis axes, while *measures* are the numerical attributes being analyzed against the different dimensions. A dimension contains members that are organized hierarchically into levels, each level having a different granularity, from coarse at the most aggregated level to fine at the most detailed level. The members of one level can be aggregated (regrouped) to form the members of the next higher level. The measures at the finest level of granularity can be aggregated or summarized following this hierarchy and provide information at the higher levels according to the aggregation rules or algorithms.

A set of measures aggregated according to a set of dimensions forms what is often called a *data cube* or *hypercube* (Thomsen et al. 1999). Inside a data cube, possible aggregations of measures on all the possible combinations of dimension members can be precomputed. This greatly increases query performance in comparison to the conventional transaction-oriented data structures found in relational and object-relational database management systems (DBMS).

The common OLAP *architecture* can be divided into three parts: the multidimensional database, the OLAP server that manages the database and carries out the different calculations, and finally the OLAP client that accesses the database via the OLAP server. This access allows the end user to explore and analyze the data using different visualization methods and adapted operators (Bédard et al. 1997) such as *drill-down* (show-details), *roll-up* (show a more global picture, also called *drill-up*), *drill-across* (show another theme at the same level of details), and *swap* (change a dimension for another one).

Finally, it is commonly found in the literature that the multidimensional approach of analysis is more in agreement with the end user's mental model of the data (Codd et al. 1993, Yougworth, 1995). Based on this approach, the interface of a tool exploring the multidimensional paradigm, such as OLAP, is usually very intuitive and the user can perform analysis ranging from simple to complex, mostly by clicking on the data being organized in a meaningful way (Yougworth 1995). This adds to the fact that the multidimensional data structure is optimized for rapid ad hoc information retrieval (OLAP Council 1995), which greatly facilitates the data exploration and analysis process.

SOLAP Analysis Technique for Spatial Data

Traditional OLAP offers good support for simultaneous usage of descriptive, temporal, and spatial dimensions in a multidimensional analysis process. Descriptive dimensions are used to describe the data to be analyzed. The temporal ones take into account the temporal aspect of the analysis and the spatial dimensions allow for the spatial reference of the phenomena under study. Using traditional OLAP tools, however, the spatial dimensions are treated like any other descriptive dimension, without consideration for the cartographic component of the data. OLAP tools present serious limitations in support of spatiotemporal analysis (no spatial visualization, practically no spatial analysis, no map-based exploration of data, etc.).

Data visualization facilitates the extraction of insight from the complexity of the spatiotemporal phenomena and processes being analyzed, as well as offering a better understanding of the structure and relationships contained within the dataset. In the context of information exploration, maps and graphics do more than make data visible; they are active instruments in the end user's thinking process (Rivest et al. 2001). Without a cartographic display, OLAP tools lack an essential feature, which could help spatiotemporal exploration and analysis. A SOLAP tool remedies this lack because it supports the geometric spatial dimensions that can include geometric shapes spatially referenced on a map. This type of spatial dimension allows the dimension members to be visualized and queried cartographically.

A SOLAP system can be defined as a visual platform built especially to support rapid and easy spatiotemporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays (Bédard et al. 2005).

OLAP-SOLAP is geared towards decision support for it is designed from the start to be easy, rapid, and multigranular.

Easy. The ease comes from the ability to conduct analysis without having to master a query language or to understand the underlying structure of the database, which may be very complex in the particular case of spatiotemporal databases (Rivest et al. 2001). In fact, the analyst interacts directly with the data and focuses on the results of the analysis rather than on the procedures required by the tool to perform the analysis process.

Rapid. It is rapid because data are preaggregated, computation time is reduced, and it is possible to provide very fast answers to complex queries. This allows the user to maintain his or her flow of thought, without his or her attention being distracted by slow response time (Rivest et al. 2001).

Multigranular. Decision support typically requires access to aggregated information (the global view) as well as more detailed information, typically following several levels of granularity of information. Such levels of granularity apply to the phenomena being analyzed as well as their time and space dimensions. Combining the thematic dimensions with the spatial and time dimensions allows one to analyze the evolution of phenomenon, to discover trends, to unveil correlations, etc.

SOLAP fundamental concepts, historical perspective, and extensive references can be found in Rivest et al. 2001, Bedard et al. 2005, Tchounikine et al. 2005, Ferri et al. 2002, while SOLAP prototypes are described by Kouba et al. 2000, Stefanovic et al. 2000, Ferreira et al. 2001, Shekhar et al. 2001, Fidalgo et al. 2004, Bedard et al. 2005, Scotch and Parmato 2005, Silva et al. 2005, among others.

Coupling Multiagent Geosimulation and OLAP–SOLAP Analysis Techniques for Better Geosimulation Output Data Analysis

Analyzing properly the simulation input and output is critical for a successful simulation study (Anu 1997, Groumpos et al. 2002, Ali and Moulin 2005). Figure 3 presents the main steps of the method we propose to develop multiagent geosimulations (Ali and Moulin 2005). In the current paper, we discuss only the input/output data analysis steps (gray boxes in Figure 3). The other steps of our approach are presented in detail in Ali and Moulin (2005) and will not be discussed in this paper.

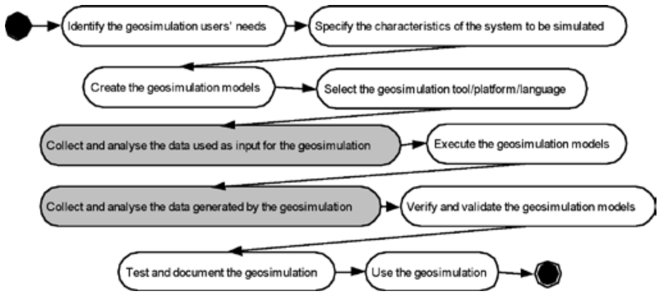


Figure 3. The main steps of the multiagent geosimulation method

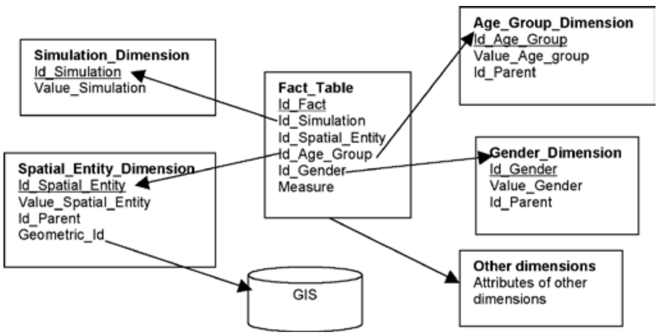


Figure 4. The structure of the database that contains the simulation output data

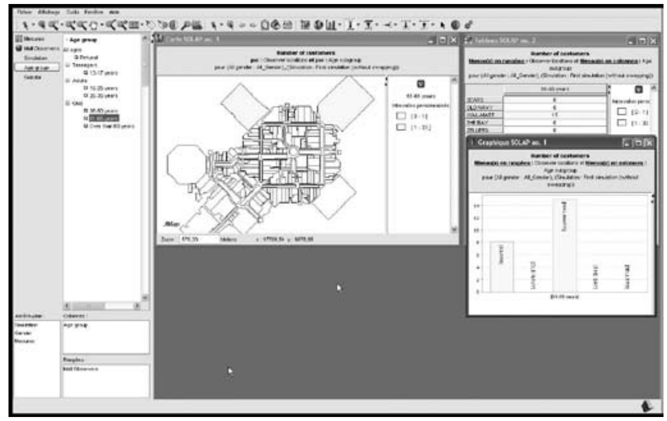


Figure 5. Distribution of software Shopper agents that are between 51 and 65 years old for five major stores

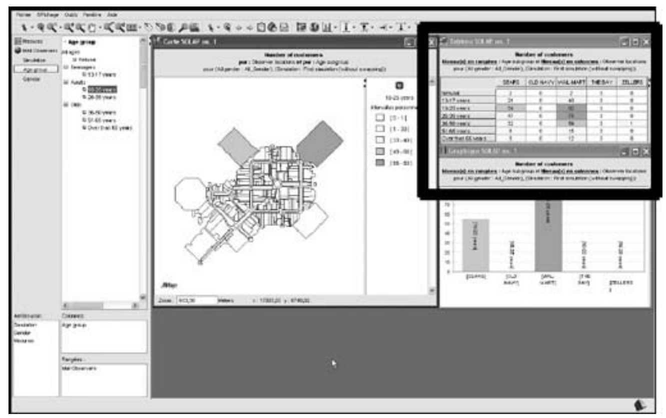


Figure 6. Distribution of software Shopper agents for five major stores. The table shows the number of Shoppers per age group for each store. The map and bar chart show the number of Shoppers that are between 18 and 25.

In our method, the simulation output data are obtained using software agents called *observer agents*. The mission of these agents is to gather data about the Shopper agents that enter their perception area. This data is recorded in a database (see Figure 4 for a simplified view of the database structure) and analyzed after the simulation completion. In this database, we can distinguish two types of tables. The *fact table,* which contains the measures that will be analyzed, and the *dimension tables,* which contain data about each hierarchy of data. The database structure contains some nonspatial dimensions (age group, gender) of the Shopper agents and a spatial dimension that contains the stores and the corridors in the mall.

The data analysis of the geosimulation output (nonspatial and spatial data) is integrated in a SOLAP extension for GIS called JMap Spatial OLAP (Bédard et al., 2005). Using this OLAP–SOLAP tool, we can analyze and explore the simulation output of the shopping behavior in the Square One mall using the nonspatial and spatial dimensions. We can visualize the distribution of the Shopper agents by store or corridor and according to the nonspatial dimensions of these agents.

For example, the user can visualize cartographically the distribution of the Shopper agents in five major stores of Square One mall (Wal-Mart, Sears, Zellers, Old Navy, and The Bay) based on the age group and the gender dimensions. Figure 5 shows the distribution of the Shopper agents that are between 51 and 65 years old for the five major stores using the SOLAP tool. In this figure we can see only the non-null distribution (8 for Sears and 15 for Wal-Mart).

If the user wants to see which categories of shoppers visit the other stores (Zellers, Old Navy, and The Bay) in terms of the age group dimension, he or she can use the SOLAP drill-up operation on the age group dimension to see the distribution for all the ages (see Figure 6). The user can see that Sears and Wal-Mart are also most visited by Shopper agents that are between 18 and 25 years old (54 for Sears and 82 for Wal-Mart). The user can also observe that the Zellers store is visited by one Shopper agent that is between 36 and 50 years old. In this figure, the white areas are not visited by this category of Shoppers (distribution is null) and colored areas correspond to the most visited ones (orange) and less visited ones (yellow).

The SOLAP tool allows the user to use the chart display to study the distribution based on all the age groups (see Figure 7).

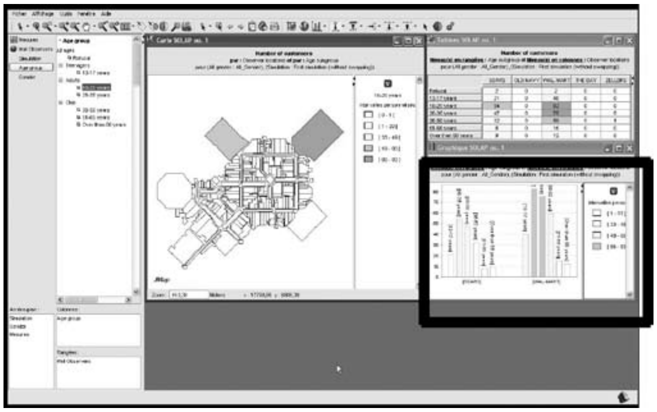


Figure 7. Distribution of software Shopper agents for five major stores. The bar chart shows the number of Shoppers per age group for Sears and Wal-Mart.

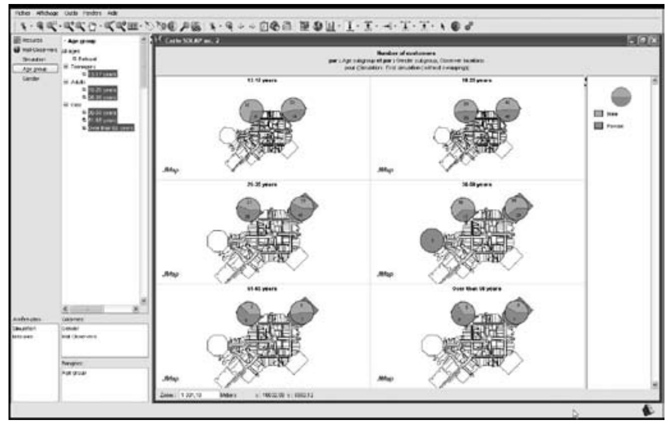


Figure 9. Cartographic representation of the distribution of software agents consisting of one map for each age group and superimposed pie charts to represent gender

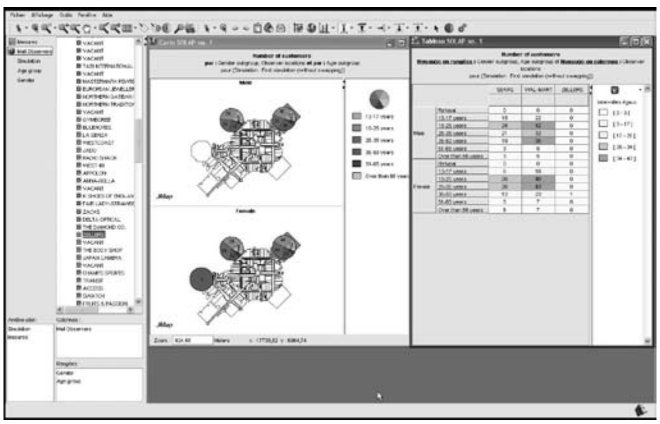


Figure 8. Cartographic representation of the distribution of Shopper agents comprising two maps (one for male and another for female agents) that include superimposed pie charts to present the distribution of each age group

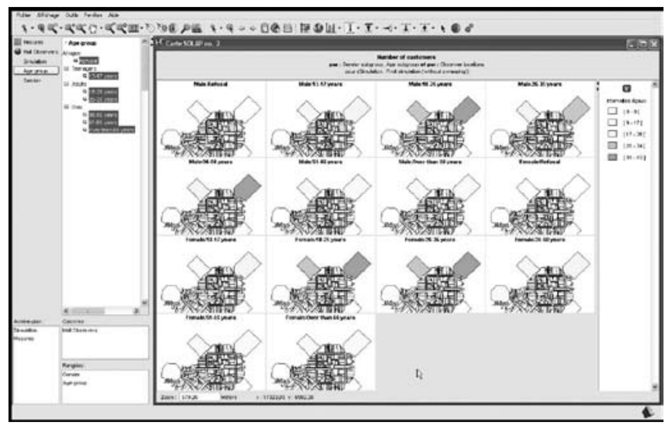


Figure 10. Cartographic representation of the distribution of software agents consisting of one map for each combination of age group and gender

In this case we focus only on the non-null distributions and we can represent the six age groups in the same bar chart.

The SOLAP tool can also present maps of the distribution for all genders and age groups. In Figure 8, we show a cartographic representation comprising two maps (one for male and another for female agents) that include superimposed pie charts to present the distribution of each age group.

It is also possible to show the same information but in another way by using multimaps (one for each age group) and pie charts to represent gender (see Figure 9).

The SOLAP tool allows the user to visualize maps of the distribution of Shopper agents that are based on combinations of members from different dimensions. Figure 10 shows 14 maps, each one representing the distribution of Shopper agents for a combination of members from the age group and the gender dimensions (seven age groups and two genders).

The SOLAP tool is very flexible in terms of the visualization and exploitation of nonspatial and spatial data. It can display data using maps, pie charts, histograms, or other visualization modes.

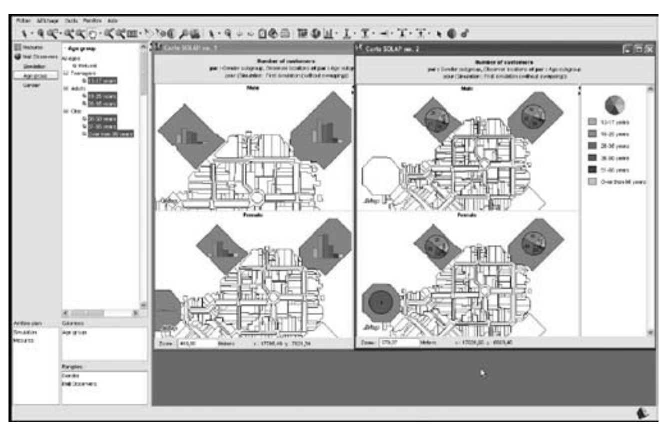


Figure 11. Cartographic representation of the distribution of software agents by age group (combining pie charts or histograms on map)

Figure 11 shows an example of visualization flexibility where the same analysis is displayed using superimposed pie charts or histograms on maps.

These results show how a user can take advantage of multidimensional analysis techniques (OLAP–SOLAP) to explore easily and rapidly both nonspatial and spatial output data generated by our multiagent geosimulation prototype. The results are directly presented on the map with different levels of detail. This easiness and quickness in the exploration and visualization of nonspatial and spatial data help the user make efficient decisions about both the nonspatial and spatial components of the simulation. This dynamic aspect allows the user to create thousands of displays (maps, tables, and diagrams) using the dataset without having to store each display individually or to learn a query language or to wait more than ten seconds for any query. These capabilities set it apart from traditional geographic information systems. All required know-how to produce maps is embedded in SOLAP technology. End users do not need to know a GIS query language, GIS functions, the database structure, and graphic symbology rules to produce state-of-the-art views on the data.

DISCUSSION AND CONCLUSION

In the literature, several researchers propose different techniques to analyze simulation outputs (Seila 1992, Kelton 1997, Alexopoulos et al. 1998, Sanchez 2001, Alexopoulos 2002). Decision makers often have difficulties interpreting these outputs because they need to manipulate the simulation results using mathematical or statistical tools that are not obvious for the average user. To overcome this shortcoming, Grier (1992) proposed a statistical analysis technique, based on tables and graphs, that can be used to analyze simulation output. This visual display of the results quickly conveys information about the model that might require hours of study to glean from mathematical and statistical analysis results. Managers who must rely on simulations to support their decisions find analysis based on tables and graphs easier to understand rather than browsing pages of numbers generated by an analytical model. Blaisdell et al. (1992) propose SIMSTAT, which is a tool for simulation analysis that is based on the graphical analysis techniques and is combined with several discrete-event simulation tools. The analysis techniques mentioned previously are efficient for event-discrete simulations or other kinds of simulations, but they present some limitations because they do not deal with spatial or geographic data. In certain kinds of simulations, such as urban simulation (US), geosimulation (GS), or multiagent geosimulation (MAGS), spatial data is an important element for decision making. In these kinds of simulations, we need more efficient ways to support spatial analysis (tables and graphs do not allow sophisticated spatial analysis and do not provide spatial visualization and map-based exploration). In this paper we showed how we can use multidimensional analysis techniques called OLAP–SOLAP to analyze geosimulation outputs. What distinguishes our approach from the techniques mentioned previously is that it profits from the advantages of the OLAP technique to present the output analysis to the users *easily* and *rapidly*. Our approach also takes into account the analysis of nonspatial and spatial data generated by multiagent geosimulations using the SOLAP analysis technique. The advantages of coupling OLAP–SOLAP with multiagent geosimulations make the output analysis more sophisticated (easiness and quickness of visualization, and the addition of cartographic presentation) and compatible with the user's mental models.

In this paper we presented how we can improve simulation output data analysis using the Online Analytical Process (OLAP) analysis technique. We also showed how we can use an extension of the OLAP technique called Spatial Online Analytical Process (SOLAP) to analyze spatial data generated by simulations that involve spatial data. Based on a multiagent geosimulation prototype, we simulated human shopping behavior in a shopping mall. With this illustration we saw how easy and efficient it is for the simulation's users to use multidimensional techniques such as OLAP–SOLAP to analyze nonspatial and spatial data generated by multiagent geosimulations.

In the future, we will apply the OLAP–SOLAP analysis techniques to analyze nonspatial and spatial data generated by other multiagent geosimulation prototypes in various application domains. This will help verify the applicability of the proposed approach in different contexts of use.

Acknowledgments

The authors are thankful to Canada GEOIDE Network of centres of excellence in geomatics, to Canada Natural Sciences and Engineering Research Council and to Canada NSERC Industrial Research Chair in Geospatial Databases for Decision-Support for their financial support. Many thanks to the Square One management who allowed our team to carry out the survey and use data. The development of the MAGS platform has been funded by the Defence (RDDC Valcartier). Many thanks to the MAGS team and especially, W. Chaker, J. Hogan, J. Perron and F. Belafkir for their work on various aspects of the MAGS platform.

About the Authors

Dr. Walid Ali obtained a Ph.D. in Computer Science from Laval University (Quebec) (supervised by Bernard Moulin since 2001). His main interest is in using multiagent systems to simulate behavior and phenomena in two-dimensional and three-dimensional geographic virtual spatial environments. He presently leads the development of innovative three-dimensional visualization-based software tools in the field of geometric morphometrics (GMM).

Dr. Bernard Moulin obtained an Engineering degree from Ecole Centrale de Lyon (France), a degree in Economics (University Lyon), and a Ph.D. (Applied Mathematics–Computer Science, University Lyon). He is now a full professor in the Computer Science Department at Laval University. His main interests are multiagent systems, agent-based geosimulation,

applications of AI techniques to geomatics, modeling, and simulation of software agents' conversations.

Dr. Yvan Bédard obtained a PhD in Civil Engineering (GIS) from the University of Maine at Orono and is full professor of GIS and Spatial Databases at Laval University Department of Geomatics Sciences since 1986. He is a member of the Centre for Research in Geomatics, where he acted as the Founding-Director, and is involved in the GEOIDE Network of Centers of Excellence. He is also associate researcher at the Centre Hospitalier Universitaire de Québec Research Centre (CHUQ), an associate professor to the Dept. of Geodesy and Geomatics Engineering at the University of New Brunswick and Holder of the Canada NSERC Industrial Research Chair in Geospatial Databases for Decision-Support. Dr. Bédard has collaborated with major government agencies and private industries in Canada and abroad, especially for the design of commercial products, strategic planning and project evaluations. He also is strongly supported by major university granting agencies. His main fields of interest are GIS, spatial database modeling and spatial on-line analytical processing.

Mrs. Marie-Josée Proulx holds a M.Sc. in Geomatics Sciences from Laval University. She works at the Centre for Research in Geomatics at Laval University as a research professional within the GIS and spatial databases team. She has contributed to several papers and conferences and given continuing education in Canada and Europe. Her professional interests include spatial databases modeling, data warehousing and metadata management.

Mrs. Sonia Rivest holds a M.Sc. in Geomatics Sciences from Laval University. She works at the Centre for Research in Geomatics at Laval University as a research professional within the GIS and spatial databases team. In addition to having experience in industry, she has contributed to several papers and conferences and given continuing education in North Africa. Her professional interests include multidimensional databases, spatial data warehouses and SOLAP.

References

Alexopoulos, C. 1998. Advanced methods for simulation output analysis. Proceedings of the 1998 Winter Simulation Conference, Washington, December 13-16, 1998.

Alexopoulos, C. 2002. Output data analysis for simulations. Proceedings of the 2002 Winter Simulation Conference, California, December 8-11, 2002.

Ali, W., and B. Moulin. 2005. Towards a generic approach to develop 2D-3D multiagent geosimulation. To appear in Proceedings of the Agent Directed Simulation (ADS '05) (Part of the 2005 Spring Simulation MultiConference (SpringSim '05)), San Diego, California, April 2-7, 2005.

Anu, M. 1997. Introduction to modeling and simulation. Proceedings of the 29th Conference on Winter Simulation, Atlanta, Georgia, December 7-10, 1997.

Bédard, Y. 1997. Spatial OLAP, vidéoconférence. 2eme forum annuel sur la R-D, Géomatique VI: un monde accessible, Montréal. November 13-14, 1997.

Bédard, Y., S. Rivest, and M. J. Proulx. 2005. Spatial online analytical processing (SOLAP): concepts, architectures and solutions from a geomatics engineering perspective. In Data warehouses and OLAP: concepts, architectures and solutions. Idea Group Publishing. Hershey, PA. On press.

Bédard, Y., M. J. Proulx, and S. Rivest. 2005. Enrichissement du OLAP pour l'analyse géographique: exemples de réalisations et différentes possibilités technologiques. Revue des nouvelles technologies de l'information, B-1, 1-20.

Benenson, I., and P. Torrens. 2003. Geographic automata systems: a new paradigm for integrating GIS and geographic simulation. In the Proceedings of the GeoComputation Conference, September, 2003.

Blaisdell, W. E., and J. Haddock. 1992. SIMSTAT: a tool for simulation analysis. Proceedings of the 1992 Winter Simulation Conference, 1992.

Bousquet, F., and I. Bakam. 1998. Proton, H., and C. Le Page. Cormas: common-pool resources and multi-agent systems. Lecture notes in artificial intelligence 1416, 826-38.

Caron, P. Y. 1998. Etude du potentiel de OLAP pour supporter l'analyse spatio-temporelle. Mémoire de M. Sc., Département des sciences géomatiques, Faculté de foresterie et géomatique, Université Laval, 1998, 132.

Codd, E. F., and C. T. Salley. 1993. Providing OLAP (on-line analytical processing) to user-analysts: an IT mandate. Hyperion white paper, 1993, 20 pp.

Dibble, C., and P. G. Feldman. 2004. The GeoGraph 3D Computational Laboratory: network and terrain landscapes for RePast. Journal of Artificial Societies and Social Simulation 7(1), 2004.

Ferreira, A. C., M. L. Campos, and A. Tanaka. 2001. An architecture for spatial and dimensional analysis integration. In the Proceedings of SCI 2001, Volume XIV, Computer Science and Engineering, Part II, 392-95.

Fidalgo, R. N., V. C. Times, J. Silva, and F. F. Souza. 2004. GeoDWFrame: a framework for guiding the design of geographical dimensional schemas. In the Proceedings of DaWaK 2004, Lecture Notes in Computer Sciences 3181, 26-37.

Gill, H. S., and P. C. Rao. 1996. The official geode to data warehousing, QUE Corporation, 1996, 382 pp.

Grier, A. D. 1992. Graphical techniques for output analysis. Proceedings of the 1992 Winter Simulation Conference, 1992.

Groumpos, P. P., and Y. Merkuryev. 2002. A methodology of discrete-event simulation of manufacturing systems: an overview. Studies in informatics and control: with emphasis on useful applications of advanced technology. Volume 11, Number 1, 103-110. March 2002.

Kelton, D. W. 1997. Statistical analysis simulation output. Proceedings of the 1997 Winter Simulation Conference, 1997.

Koch, A. 2001. Linking multi-agent systems and GIS—modeling and simulating spatial interactions. Department of Geography RWTH Aachen. Angewandte Geographische Informationsverarbeitung XII, Beiträge zum AGIT-Symposium Salzburg 2000, Hrsg.: Strobl/Blaschke/Griesebner, Heidelberg, 2001, 252-62.

Kouba, Z., K. Matousek, and P. Miksovsky. 2000. On data warehouse and GIS integration. In the Proceedings of DEXA 2000, Lecture Notes in Computer Sciences 1873, 604-13.

Mandl, P. 2000. GeoSimulation—experimentieren und problemlösen mit GIS-modellen. In Angewandte Geographische Informationsverarbeitung XII, Hrsg.:Strobl/Blaschke/Griesebner, Wichmann Heidelberg, 2000, 345-56.

Maslow, P. 1970. Motivation and personality, 2nd Ed. NY: Harper and Row, 1970.

Moulin, B., W. Chaker, J. Perron, et al. 2003. MAGS project: multi-agent geosimulation and crowd simulation. In the Proceedings of the COSIT '03 Conference, Ittingen (Switzerland). Kuhn, Worboys, and Timpf, Eds., Spatial information theory. Springer Verlag LNCS 2825, 2003, 151-68.

Najlis, R., and M. North. 2004. Repast for GIS. Paper presented in the Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity, and Emergence, Chicago, IL, 2004.

Pendse, N. 2000. Glossary, the OLAP report, http://www.olapreport.com/fasmi.htin.

Perron, J., and B. Moulin. 2004. Un modèle de mémoire dans un système multi-agent de géo-simulation. Revue d'intelligence artificielle. Hermes, 2004.

Rivest, S., Y. Bédard, and P. Marchand. 2001. Toward better support for spatial decision making: defining the characteristics of spatial on-line analytical processing (SOLAP). Geomatica 55(4): 539-55.

Rivest, S., P. Gignac, J. Charron, and Y. Bédard. 2004. Développement d'un système d'exploration spatio-temporelle interactive des données de la Banque d'information corporative du ministère des Transports du Québec. Proceedings, Geomatics 2004, Conference of the Canadian Institute of Geomatics—Montreal Section, October 28-29, 2004.

Sanchez, S. M. 2001. ABC's of output analysis. Proceedings of the 2001 Winter Simulation Conference, Arlington, December 9-12, 2001.

Scotch, M., and B. Parmanto. 2005. SOVAT: Spatial OLAP visualization and analysis tool. In the Proceedings of the 38th Hawaii International Conference on System Sciences, 2005, 142.2.

Shekhar, S., C. T. Lu, X. Tan, et al. 2001. Map cube: a visualization tool for spatial data warehouses. In Miller, H., and J. Han, Eds., Geographic data mining and knowledge discovery. London: Taylor and Francis, 2001, 74-109.

Seila, A. F. 1992. Advanced output analysis for simulation. Proceedings of the 1992 Winter Simulation Conference, Arlington, December 13-16, 1992.

Silva, J., V. Times, R. Fidalgo, and R. Barros. 2005. Providing geographic-multidimensional decision support over the Web. In APWeb 2005: 7th Asia-Pacific Web Conference, Lecture Notes in Computer Sciences 3399, 2005, 477-88.

Stefanovic, N., J. Han, and K. Koperski. 2000. Object-based selective materialization for efficient implementation of spatial data cubes. IEEE Transactions on Knowledge Discovery and Data Engineering 12(6): 938-58.

Tchounikine, A., M. Miquel, R. Laurini, et al. 2005. Panorama de travaux autour de l'intégration de données spatio-temporelles dans les hypercubes. Revue des nouvelles technologies de l'information, B-1, 2005, 21-33.

Thomsen, E., G. Spofford, and D. Chase. 1999. Microsoft OLAP solutions. NJ: John Wiley and Sons, 1999, 495 pp.

Yougworth, P. 1995. OLAP spells success for users and developers. Data based advisor, 1995, 38-49.

An Interoperable Portal Supporting Prototyping Geospatial Applications

Myra Bambacus, Phil Yang, John Evans, Marge Cole, Nadine Alameh, and Stephen Marley

Abstract: *Earth observations and simulation results can be integrated with decision-support tools to support geospatial applications of national and international interest. Traditional methods of application development through tightly coupled components can no longer satisfy increased demands and urgently needed geospatial applications, such as disaster management. We examine the life cycle of such integrated applications, and utilize an interoperable prototype known as the Earth Science Gateway (ESG) to share earth observations and earth science simulations, such as global precipitation, to support the prototyping needs of such geospatial applications. Service-Oriented Architecture and spatial Web services are utilized to design and develop the ESG, which is illustrated in facilitating and accelerating the development of geospatial applications by leveraging Earth observations, simulation models, and decision-support tools.*

INTRODUCTION

NASA's Earth Observing System (EOS, including satellite sensors and satellite data-receiving systems) is generating more than two terabytes of geoscience data daily, and has archived more than four petabytes of data in its Distributed Active Archive Centers (DAACs). Among these datasets, about 34 million data products and 640 TB data were disseminated to more than 2 million distinct users in 2004 (NASA 2005). To further improve the usage of the data, NASA and its partner agencies identified 12 national applications (Birk et al 2006). These application areas integrate the earth observations, earth system modeling, and decision support tools to support national and societal needs, such as public health and coastal management. Integration of global earth observation data and earth system models in applications from regional to global benefit also helps to build the Global Earth Observation System of Systems (GEOSS, GEO 2005).

The integration processes, for national applications and GEOSS applications, are illustrated in Figure 1:

1) Earth observation systems acquire data through remote sensors and in situ sensors.
2) The earth observation data are fed into earth system models or the decision-support tools.
3) The earth system modeling results are fed into decision-support tools.
4) The decision-support tools outputs are used for policy and management decisions.
5) The feedback from the policy and management decisions are used to improve earth observation systems and earth system models.

Within the last few decades, geospatial information (such as observation data and simulation outputs) has been used widely for decision-supporting applications from a global level, such as the crop yield predictions for diplomatic use (Doraiswamy 2004), to a local level, such as the West Nile Virus surveillance (Gutro 2002). Each of the components involved in the applications is a valuable asset. With the objective to develop national applications and GEOSS applications by integrating these assets, it becomes important to share these assets in an interoperable and fast manner. Services provide an opportunity to achieve this requirement. For example, the geospatial components (see Figure 1) can be extracted and developed as different services:

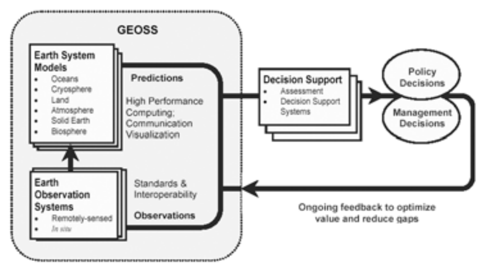


Figure 1. GEOSS architecture diagram (GEO 2005)

1) The earth observation systems can be enhanced to provide interoperable data services.
2) The earth system models can provide simulation output as interoperable data or information services.
3) The decision-support tools can provide geospatial decision-support processing services.
4) The decision makers and practitioners can observe the impacts of decisions made and provide the impacts as a quality of service (QoS) feedback to optimize or improve the services involved in the process.

Many data and information services have been developed to share earth observations (such as the EOS Data Gateway where users can search and download or order data, EDG, 2006) and model simulations (such as the Global Modeling and Assimilation Office, which is responsible for global atmospheric phenomena simulations within NASA, GMAO, 2003). Online catalogs, such as the Federal Geographic Data Committee (FGDC) Clearinghouse (FGDC 1996), NASA's Global Climate Master Directory (GCMD, http://gcmd.gsfc.nasa.gov/), and NASA's EOS Clearinghouse (ECHO, http://www.echo.eos.nasa.gov/) have been developed to facilitate the discovery of the services. The services and catalogs provide interoperable accessibility to existing earth observations and model simulations, such as the Global Mosaic WMS service based on TM images (http://onearth.jpl.nasa.gov/). Geospatial interoperability can help to leverage and link these catalogs, and to chain services (Alameh 2003) together within a larger Service-Oriented Architecture (SOA, W3C 2003). Therefore, the services can be built once and used many times to increase the return on investments and reduce time for prototyping through interoperability (Bambacus and Reichardt 2006).

Recent developments in Web portal technology provide a possible interoperable platform to share the services (Goodchild 2006) and to support application prototyping. Observing these developments, the federation of Earth Science Information Partners (ESIP, http://www.esipfed.org/) is identifying different users' information requirements and technology needs (as shown in Table 1) for sharing geospatial resources (http://wiki.esipfed.org/) via the Earth Information Exchange (ESIP 2005).

Observing the requirements, we investigated using the ESG as an interoperable portal to quickly prototype geospatial applications. This paper reports our research and development in utilizing the Web services (Deitel et al 2002) and spatial Web portal (Yang et al 2007) to leverage legacy geospatial components.

Table 1. Information Sharing Requirement (Courtesy of ESIP Federation)

Users	**Information Required**	**Technology**
Data Users	Ready access to all existing metadata, which describe the geospatial datasets	Catalog and interoperable data viewing portal access
Researchers	Data in scientific data formats compatible with common scientific data models (e.g., netCDF, HDF-EOS)	Sophisticated portal access
Policy makers and general public	Advanced data products (analysis tools, models, simulations, decision support products)	Portal access to specific decision support applications
Educators	Educational products (collections, simulations, informational videos, lesson plans)	Portal access to populated educational communities

The following sections introduce the technical requirements of a geospatial interoperable portal, an architecture that supports those requirements, and the design of ESG using such architecture. The last two sections provide an example of how the ESG can be used to prototype applications and include conclusions and discussions.

Table 2. Characteristics of a Geospatial Portal (Adopted from Rose 2004)

Characteristics	**Explanation**	**Relevant SOA and Portal components**
Interoperable	A portal should be able to access other portals and legacy systems through interoperable interfaces, such as Web services, at different levels, such as metadata, data, and services.	Service chaining (Alameh 2003)
Compliant with geospatial standards and specifications	A geoscience portal should support geospatial specifications, such as OGC Web Services, and standards, such as ISO/TC211 standards, so that relevant geospatial applications can be accessed in consistent, predictable, and interchangeable manners.	OGC Web Services and FGDC/ISO standards
Vendor neutral	Legacy systems were developed based on different vendors' solutions. An interoperable portal should comply with open community standards and be independent of vendor-specific requirements.	Service discovery and service chaining
Scalable and expandable	A portal should comply with the general software engineering requirements on component reuse and expansion. For example, JSR 168 portlets' architecture and specifications can be adopted to facilitate the portal scalability and expandability.	Publish-find-bind, JSR 168
Web services	Portal should comply with Web-interfacing standards, such as HTTP for communication, XML for encoding requests and responses.	OGC Web Services
End-user applications	A portal for geoscience should support end-user applications.	Clients/ applications

GEOSPATIAL NEEDS FOR AN INTEROPERABLE PORTAL

To support the user requirements, an interoperable portal should be able to provide support to specific geospatial characteristics (see Table 2).

Legacy earth science components are diverse and heterogeneous because various providers have developed these components to satisfy different requirements. Sharing these components across the earth observation community presents many challenges and requires the incorporation of different levels of standards, such as community-based standards (OpeNDAP, http://opendap.org/, and NetCDF, http://www.unidata.ucar.edu/software/netcdf/, or the HDF–EOS, http://hdfeos.net/, data format) and broader international standards. A series of standards have been developed within the frameworks of the FGDC, the Open Geospatial Consortium (OGC), and the International Standards Organization/Technical Committee 211 standards (ISO/TC211).

Among the standards and specifications developed by the FGDC, OGC, and ISO/TC211, significant specifications for interoperable portals include the Catalog Service for the Web (CS-W, Nebert and Whiteside 2004), Web Map Service (WMS, de La Beaujardiere 2004), Web Feature Service (WFS, Vretanos 2002), Web Coverage Service (WCS, Evans 2003), Geography Markup Language (GML, Cox 2004), and Styled Layer Descriptor (SLD, Müller and MacGill 2005). CS-W and Z39.50 (ANSI/NISO 2003) specifications facilitate communication among catalogs. WMS is for distributed visualization of geospatial information as mapped images. WCS facilitates the sharing of data from earth observations, scientific assimilations, or modeling. WFS helps in sharing discrete data objects, and SLD supports rendering data obtained via WFS or WCS; while such rendering services are often denoted Web Feature (or Coverage) Portrayal Services (WFPS, WCPS—cf. Lansing 2002).

Based on these service specifications and standards, the SOA can provide a publish-find-bind pattern to support sharing components and prototyping applications (as illustrated in Figure 2): (1) service providers publish service descriptions to a catalog; (2) clients find published services through service descriptions; (3) services are bound according to application logic to support specific client applications.

EARTH SCIENCE GATEWAY CONCEPTUAL ARCHITECTURE

Based on the interoperability and Web portal technology described previously, we developed a conceptual architecture (shown in Figure 3) to share geospatial components and to support application prototyping. The ESG was implemented with a conceptual model in mind, in which future and legacy components (yellow boxes) are shared as services, while the ESG Catalog acts as the registry and index for the services. ESG clients can discover needed components from the catalog and bind them to form applications.

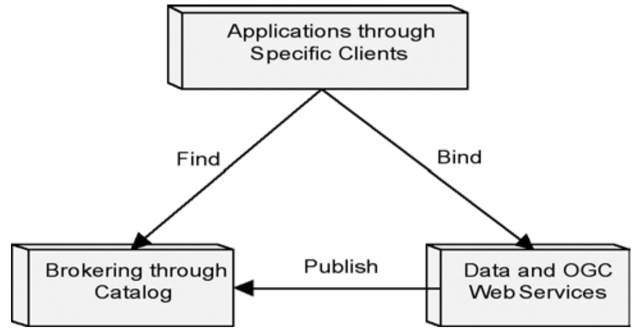


Figure 2. Service-oriented architecture (W3C 2003)

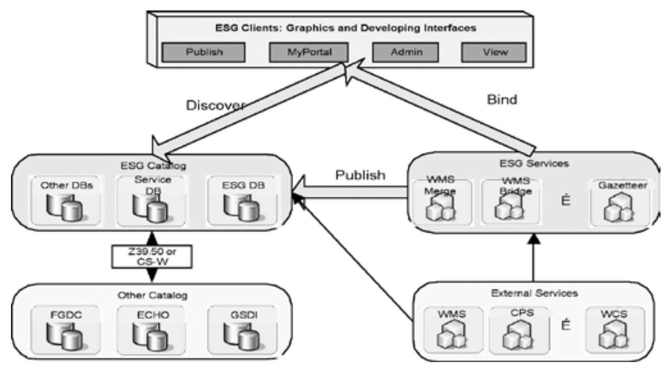


Figure 3. The ESG logic architecture

The *External Services* box provides data or information services via WMS and WCS. For example, NASA's Jet Propulsion Laboratory provides WMS access to its geophysical observation data and simulations, such as ocean color or sea surface temperature (JPL 2006). The *Other Catalog* box provides catalogs, such as the NSDI Clearinghouse, for searching geospatial data, information, and services.

The *ESG Services* box provides a set of facilitating services, such as WMSBridge, WMSMerge, and Gazetteer. WMSBridge translates requests and responses between WMS and legacy map services (such as ArcIMS, a proprietary sevice produced by ESRI). The WMSMerge service reprojects the outputs of multiple WMSes onto a common coordinate reference system and overlays them onto one image. The Gazetteer service translates place-names (cities, rivers, mountains, etc.) into precise geospatial coordinates.

Finally, the *ESG Catalog* box is populated with service metadata of external services and ESG services. The ESG Catalog can also connect to other catalogs through Z39.50 and CS-W to share metadata registered in other catalogs.

The *ESG Clients,* which is a pure html/Javascript-based Web page that will be loaded to a Web browser when being accessed, provides visualization, searching, publishing, administration, harvesting, and personalization of prototype applications, such as air-quality applications, in a service-chaining manner.

EARTH SCIENCE GATEWAY TECHNICAL WORKFLOWS AND CLIENTS

The ESG was implemented as a loosely coupled service architecture, where services can be chained through interoperable interfaces to support different clients (see Figure 4). The essential client-support prototypes include *publisher*, *discovery*, and *viewer*. Besides OGC Web Services, the ESG includes some functional services to facilitate the sharing of existing services, in particular WMSMerge, WMSBridge, and Gazetteer service.

Publisher

The *publisher* can be employed to register services and *Harvester* can help to harvest service by connecting to the *catalog* (see Figure 4). The publisher supports the geographic content standard ISO 19115 (ISO 2003) and the FGDC (1998) metadata standard. The system can populate its own records from hypertext links to ISO 19115-structured metadata with the capabilities of a service. Users can interactively input metadata for their services when publishing through the publishing wizard (shown in Figure 5).

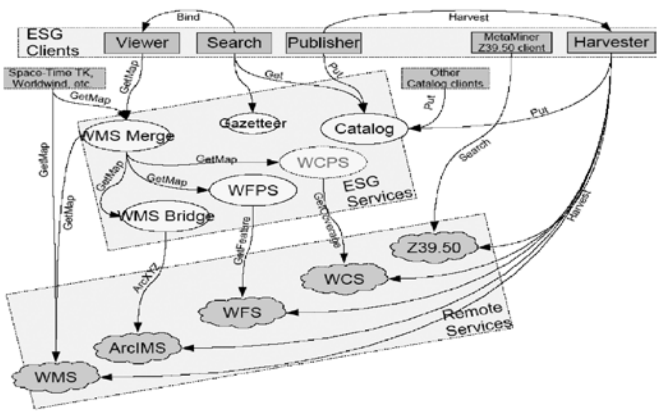


Figure 4. Service chaining–supported ESG clients

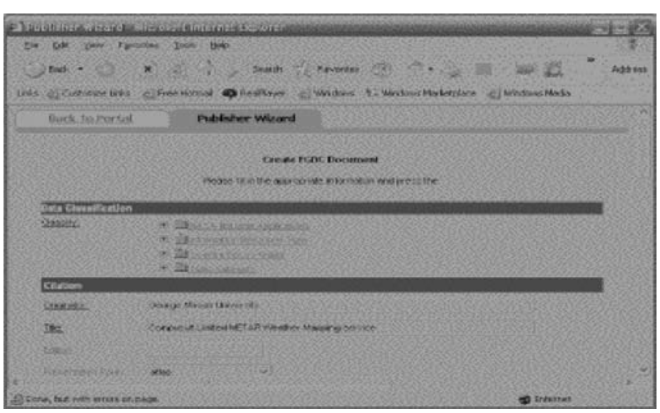


Figure 5. The publishing of a WMS to the ESG catalog will record the service metadata and basic service content for discovery purposes.

Discovery and Searching

The *search* function can locate a place through *Gazetteer*, or find services or other resources through the *user interface* (see Figure 6): services found can be *bound* into the *viewer* as depicted in Figure 7. ESG's use of the FGDC Metadata standard also allows it to search across the NSDI collection of heterogeneous catalogs.

Viewer (2-D, 3-D, 4-D)

The *viewer* can help build an application through *WMSMerge* to issue a *GetMap* request. The *GetMap* can get layers by issuing (a) *GetMap* to *WMS,* (b) *WMS Bridge* for *ArcIMS*, (c) *WFPS* for *WFS*, and (d) *WCPS* for *WCS* (see Figure 4). The ESG uses WMS to visualize the data within a two-dimensional client. The ESG is currently being extended to support visualization in 3-D (such as height) and 4-D (temporal variation) through other clients such as NASA's WorldWind (http://worldwind.arc.nasa.gov/index.html), Space Time Toolkit (STT, http://vast.uah.edu/SpaceTimeToolkit/), and Google Earth (http://earth.google.com/).

To facilitate application prototyping and system administration, the ESG also provides several utility functions (Compusult 2005), such as *Place Name Searching*, *Administration*, and *User personalization (My Portal)*.

Prototyping Applications through the Earth Science Gateway

Leveraging ESG services and the services published to the ESG catalog, ESG clients can assist in prototyping applications; furthermore, applications can be saved locally or onto the server. To revise or improve the application, another service can always be found and added to the product.

The following example is used for prototyping a wind-power locating application for the GEOSS demo in May, 2006, Beijing (CGC 2006). NASA's GMAO and NOAA's National Centers for Environmental Prediction (NCEP) collaborated on modeling and predicting earth science phenomena, such as global atmosphere circulation and wind speed. Other countries and agencies, such as Environment Canada, also developed similar earth science simulations. After the simulation results are produced, they are put into WCS and WMS and the services are registered into different

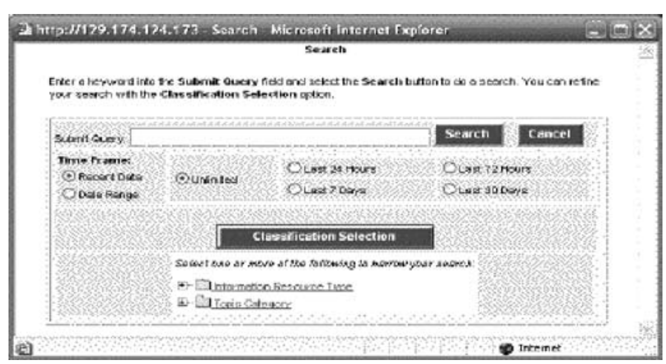


Figure 6. Search a place, data, and service through the search interface.

catalogs, such as the ESG. These global earth observations and simulations have potential to be used for national applications or GEOSS applications of societal benefits, such as energy management. For example, we can use ESG to rapidly prototype an application to identify locations for building wind farms to produce electricity in the Hainan province of China (see Figure 7).

The following workflow illustrates how to discover and integrate current wind-power information to prototype a wind-power application for identifying a wind-farm location.

(1) Search ESG (see Figure 6) using “wind” to find services of interest from tens of wind-related services, such as G5FCST Wind Shear (shown in Figure 8).

(2) Add the *G5FCST Wind Shear* service that was found to the viewer to get the desired application (see Figure 9).

(3) The application shows some wind situations but no detailed wind-speed information. Other services, such as the Mean Wind Speed service with rough (see Figure 10) and fine (see Figure 11) resolutions are then added.

This application can then be saved, and whenever a user brings up the application, updated observations and simulations will be integrated.

Through this process, an application can be prototyped quickly with the support of the ESG. In the process, users do not have to know who provided the observations or simulations, or who provided the external WMS service. Professional users only focus on the application logic by searching available services and selecting needed services. The public users only need to bring up the application through an Internet hyperlink prepared by a professional. In this example, the legacy system of collected observation data and simulations of wind speed are leveraged in the find-and-bind process.

The ESG is demonstrated in this example for its generic find-and-bind for rapid prototyping applications. However, it has the capability to bring up 3-D and 4-D visualizations and is targeted to serve the communities in sharing global earth observation data and simulations. Therefore, the ESG better serves the purpose of rapidly prototyping national applications and GEOSS applications than do other generic portals. The ESG components, such as the Client, can also be easily reused and integrated within other portals. For example, Figure 12 illustrates that the ESG client is used to support the Earth Information Exchange (http://eie.cos.gmu.edu/), a portal for exchanging geospatial information.

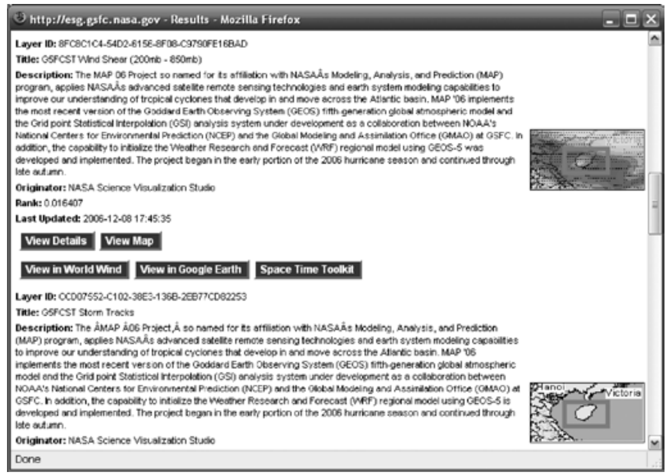


Figure 8. Search wind to find relevant services.

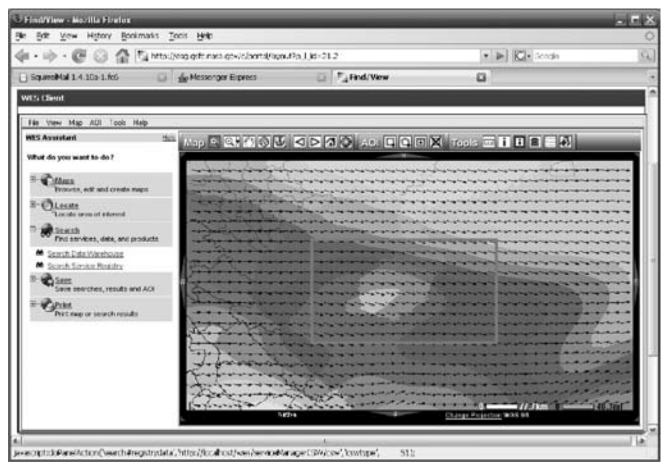


Figure 9. Wind status shown around and within Hainan province.

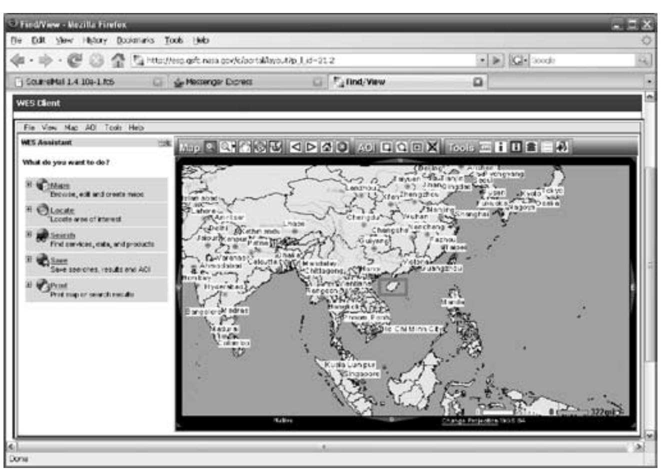


Figure 7. The pink rectangle on the map encloses the Hainan province of China.

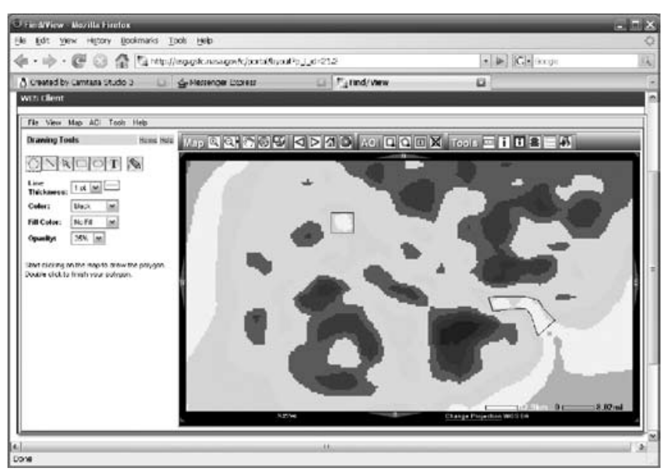


Figure 10. Wind speed can be integrated to identify the two polygons as rough areas.

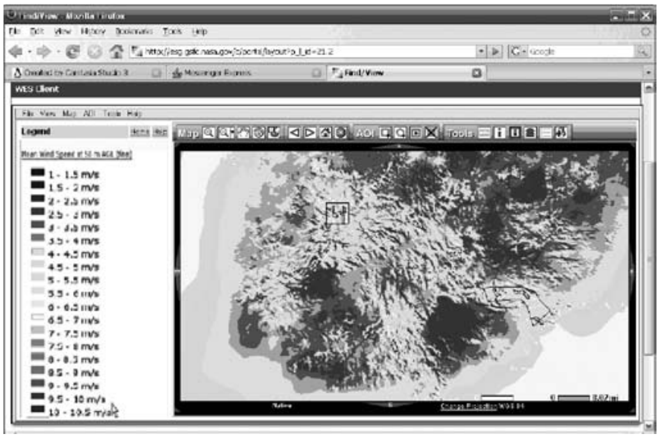


Figure 11. Finer wind speed can be integrated to identify more specific locations as the polygons within the two larger ones and three more locations are identified outside the two larger polygons.

Figure 12. The ESG client is reused and integrated into the EIE.

CONCLUSION AND DISCUSSION

This paper introduces ESG, a geospatial Web portal designed and developed to leverage the advantages of interoperability, SOA, and OGC Web Services to support prototyping applications and to reuse data by sharing earth observations, earth system modeling, and decision-support tools. In particular, ESG's interoperability provides quick and easy integration of systems and components through open interfaces for rapid prototyping (Birk et al 2006).

As a facilitating portal, the ESG provides a mechanism to prototype applications and reuse geospatial components. However, services found through the ESG discovery functions are limited by the availability and the quality of service (QoS), which depends on several factors, such as (1) the accuracy of observed data; (2) the quality of the simulation models; (3) the quality of postprocessing of information to provide the service; and (4) the reliability of the service. To support national applications prototyping, we are also evaluating, verifying, validating, and benchmarking the prototyping process with NASA partner agencies, such as the U.S. Environmental Protection Agency (EPA), which hosts national applications and GEOSS applications, such as AirNow (Dickerson and White 2006).

Acknowledgments

This paper is supported through NASA grant # NNX07AD99G.

About the Authors

Myra Bambacus is the Program Manager of NASA's Geoscience Interoperability Office (GIO).

Phil Yang received his Ph.D. in Geographic Information Science from Peking University, China, in 2000. He is an assistant professor at George Mason University, and his research is focused on geospatial information modeling, computing, and visualization.

Corresponding Address:
College of Science
George Mason University
4400 University Drive
Fairfax, VA 22030
cyang3@gmu.edu

John Evans received his Ph.D. in Computing Science from M.I.T. in 1997. He is a senior computer scientist at NASA's Geoscience Interoperability Office, and his research is focused on geospatial interoperability.

Marge Cole is a project manager at NASA's Geoscience Interoperability Office and is focused on geospatial interoperability.

Nadine Alameh received her Ph.D. in Computer Information Systems from M.I.T in 2001. She is a senior interoperability architect at NASA's Geoscience Interoperability Office, and her research is focused on geospatial Web services and interoperable architectures.

Stephen Marley received his Ph.D. from the University of College–London. His research is focused on geospatial interoperability and enterprise architecture.

References

Alameh N. 2003. Chaining geographic information Web services. IEEE Internet Computing 6(18): 22-29.

ANSI/NISO. 2003. Z39.50-2003 Information retrieval: application service definition & protocol specification. National Information Standards Organization, 2003, http://www.niso.org/standards/resources/Z39-50-2003.pdf.

Bambacus, M., and M. Reichardt. 2006. Invest in interoperability. Geospatial Solutions, February, 2006, 26-30.

Birk, R., M. Frederick, L. C. Dewayne, and M. W. Lapenta. 2006. NASA's applied sciences program: transforming research results into operational success. Earth Imaging Journal 3(3): 18-23.

CGC. 2006. Implementing the GEOSS architecture using open standards: GEOSS demonstrator using OGC specifications. The user and GEOSS architecture workshop, 22-23 May 2006, Beijing, http://geoapp2.nottingham.ac.uk:8080/GEOSS_Beijing_Demo-2006(flash)/GEOSS_Beijing_Demo-2006(flash).html.

Compusult. 2005. Web enterprise suite technical architecture, V.3.0.

Cox, S., Ed. 2004. OGC geography markup language, implementation specification, http://www.opengis.net/gml/.

de La Beaujardiere, J., Ed. 2004. Web map service ver.1.3, OGC implementation specification, http://portal.opengis.org/files/?artifact_id=5316.

Dietel H., P. Deitel, B. DuWaldt, and L. Trees. 2002. Web services: a technical introduction. NJ: Prentice Hall PTR.

Dickerson, P., and J. White. 2006. Airnow Gateway, http://www.epa.gov/scienceforum/2006/poster_abstracts/global_challenges/GC_White.pdf.

Doraiswamy, P. C., J. L. Hatfield, et al. 2004. Crop condition and yield simulations using Landsat and MODIS imagery. Remote Sensing of Environment. 92:548-59.

EDG. 2006. Earth observing system data gateway, http://delenn.gsfc.nasa.gov/~imswww/pub/imswelcome/.

Evans, J., Ed. 2003. Web coverage service Ver. 1.0, OGC implementation specification, http://portal.opengeospatial.org/files/?artifact_id=3837&version=2.

ESIP. 2005. Earth Information Exchange, http://www.esipfed.org/business/library/meetings/16th_fed_meeting/index.html.

FGDC. 1998. Content standard for digital geospatial metadata, http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf.

FGDC. 1996. FGDC Clearinghouse, http://www.fgdc.gov/clearinghouse/.

GEO. 2005. Group on earth observations, http://www.earthobservations.org/index.html.

GMAO. 2003. Global modeling and assimilation office, http://gmao.gsfc.nasa.gov/overview.php.

Goodchild, M. F., D. M. Johnston, et al. 2006. Advances in distributed and mobile computing. In McMaster R., and E. L. Usery, Eds., A research agenda for geographic information science. Philadelphia: Taylor & Francis, CRC Press, 2004, Chapter 9, http://www.ucgis.org/priorities/research/2006research/chapter_9_update.pdf.

Gutro R. 2002. Pennsylvania's West Nile Virus surveillance system gets an assist from NASA data, June 16, 2007, http://www.gsfc.nasa.gov/topstory/2002/20020830healthalliance.html.

ISO. 2003. International standard: geographic information—metadata, http://www.ncits.org/ref-docs/FDIS_19115.pdf.

JPL. 2006. NASA JPL WMS, http://wms.jpl.nasa.gov/.

Lansing, J., Ed. 2002. Web coverage portrayal service Ver.0.0.2, http://portal.opengeospatial.org/files/?artifact_id=1121.

Müller, M., and J. MacGill. 2005. Styled layer descriptor application profile of the Web Map Service: draft implementation specification, http://www.opengeospatial.org/docs/02-070.pdf.

NASA. 2005. NASA's earth system science data resources, June, 2005, pp. 1-1, http://science.hq.nasa.gov/research/earth_science.html.

Nebert, D., and A.Whiteside, Eds. 2005. Catalog services Ver.2, OGC implementation specification, http://portal.opengis.org/files/?artifact_id=5929.

Rose, L., Ed. 2004. Geospatial portal reference architecture, a community guide to implementing standards-based geospatial portals Ver.0.2, https://portal.opengeospatial.org/files/?artifact_id=6669.

Vretanos, P. A., Ed. 2002. Web feature service Ver. 1.0, http://portal.opengeospatial.org/files/?artifact_id=7176.

W3C. 2003. Web services and service oriented architecture, http://www.w3.org/2003/Talks/1211-xml2003-wssoa/.

Yang, C., J. Evans, et al. 2007. The emerging concepts and applications of the spacial Web portal. PE&RS 73(6): 691-98.

Data Transfer Solutions®

Worlds Unite

We separate ourselves from the competition by encompassing an uncommonly wide range of skills. GIS meets Public Involvement skills. Asset Management meets Web Design. Video Production merges with Transportation. The diversity of our team enables us to offer solutions that encompass the entire depth, breadth and vision of your project.

Web & Database Applications • Geographic Information Systems

Transportation • Video & Multimedia • Asset Management

Corporate Headquarters: 4037 Avalon Park Boulevard East • Orlando, Florida • 32828 • 407.382.5222

Regional Offices:

Tampa, Florida • Jacksonville, Florida • Naples, Florida • Charlotte, North Carolina

St. Louis, Missouri • San Antonio, Texas • Denver, Colorado • Fort Collins, Colorado

www.edats.com

A Comprehensive Process for Linear Referencing

Kevin M. Curtin, Greta Nicoara, and Rumana Reaz Arifin

Abstract: *This paper identifies and analyzes linear referencing as a spatial process. This is a significant departure from the existing research that defines linear referencing as a set of objects. The critical issues in the development of such a process are identified, including the determination of network representations and topology, the route structure, the determination of measures along linear features, the creation of event data, and the display and analysis of those events. A model is presented that delineates this process for streamlining the implementation of linear referencing, and providing a structure that can manage the increasing level of complexity in spatial data.*

INTRODUCTION

The primary objective of this paper is to identify, demonstrate, and analyze the necessary and sufficient requirements for exploiting linear referencing. A linear referencing process is developed and presented that expands on the extant linear referencing data models, methods, and systems that have appeared in the literature. This process is intended to provide a framework for the implementation of linear referencing among an expanding group of Geographic Information System (GIS) users.

Linear referencing can be used by many organizations, industries, and institutions that work with linear features, such as road-management organizations, transit organizations, oil and gas exploration industries, and water-resources managers, to name only a few. The common element among these industries is their use of linear features and the need to reference a position or measure along those features.

As GIS becomes more prevalent among an increasingly diverse and rapidly growing set of users, including small to midsize municipalities, government agencies at every level, and private businesses, there is an increasing demand for more sophisticated approaches to data management. When the network databases that have long been modeled in GIS (Curtin 2007a) are an important element of the analyses undertaken by these groups, the need to successfully implement linear referencing becomes an important issue, and a process for linear referencing is essential.

This research presents a comprehensive process for linear referencing. In the following section, linear referencing is formally defined and its advantages outlined. This is followed by a comprehensive literature review that discusses both the applied use of linear referencing—particularly in GIS—and the theoretical models and methods that have been developed. Based on this review, a seven-step process for linear referencing is presented and its use is demonstrated through a case study of the city of Richardson, Texas. Conclusions regarding the potential use of this process and opportunities for future research are discussed.

Linear Referencing Defined

The term *linear referencing* emerged from engineering applications where it was preferable to locate a point along a linear feature (often roads) by referencing that location to some other well-defined location, rather than using classical geographic coordinate systems. The most familiar illustration of linear referencing is the mile markers along U.S. highways (Federal Highway Administration 2001, Federal Transit Administration 2003).

Determining locations with linear referencing differs from traditional geographic coordinate and reference systems (latitude-longitude, Universal Transverse Mercator (UTM), state plane, etc.) in that the underlying entity used as a basis for measurement is not the earth, but is rather a linear feature or a set of linear features organized into a network. Just as there are myriad coordinate systems for the globe, there are multiple linear referencing systems. A common definition for a linear referencing system (LRS) is a support system for the storage and maintenance of information on events that occur along (or within) a transportation network. In this context, an LRS consists of an underlying transportation network that supplies the geographic backbone for the location of events, a set of objects with well-defined geographic locations (also known as a datum), one or more linear referencing methods (LRMs), and a set—or sets—of points or linear events that should be referenced to the underlying network. This paper will demonstrate that in many cases the underlying network and the datum are one and the same.

An LRM can be defined as a mechanism for finding and stating the location of an unknown point along a network by referencing it to a known point (Vonderohe, Chou et al. 1997). More specifically, an LRM is a process for determining a previously unknown location based on (1) a defined path along the underlying transportation network, (2) a distance along that path measured from a known datum location, and (3) optionally an offset from the path. There are several different common types of LRMs that differ based on the parts of the network used for referencing and the ways in which measures and offsets are calculated (Nyerges 1990).

Applications and Benefits of Linear Referencing

Linear referencing can be applied to any network-based phenomenon. Given the historical development of the technique of linear referencing, however, transportation applications dominate the literature. Some of the more common transportation uses are the mapping of accident, traffic stop, or other incident locations, and asset-management functions such as the recording of pavement conditions or the location of street signs, streetlights, bridges, or other traffic-related objects. Despite the historic concentration on transportation applications, significant benefits result from using linear referencing for applications in many different fields. In hydrologic modeling, linear referencing can be used to locate flow gauges along rivers or monitoring stations along creeks or pipelines. In utility facilities management, linear referencing can be used to model and display the attributes of the distribution network.

For *any* network application, using linear referencing has several primary benefits. First, locations specified with linear referencing can be readily recovered in the field and are generally more intuitive than locations specified with traditional coordinates. Second, linear referencing removes the requirement of a highly segmented linear network based on differences in attribute values. More specifically, many network attributes do not begin, end, or change values at the same points where the network is segmented: *i.e.,* speed limits do not always change at intersections, pavement quality can change at any point along a road, and stream widths can change at many different points along a stream channel. If the changes in the values of all network attributes were used to segment the network so that each segment could have a unique attribute value, this would result in an increasingly segmented (and therefore larger) database. The implementation of linear referencing allows an organization to maintain a network database with many different attribute events associated with a single, reasonably small set of network features. The implementation of linear referencing thus reduces the redundancy and potential error within the database, and it facilitates multiple cartographic representations of network attribute data.

LITERATURE REVIEW

The literature pertinent to this research falls broadly into two areas: the theoretical data models for linear referencing that have been developed and the implementation of linear referencing in GIS.

Theoretical Linear Referencing Data Models

Those who wish to apply the principles of linear referencing within GIS face a daunting set of theoretical linear referencing data models. Perhaps most significant among these are the models developed under the auspices of the National Cooperative Highway Research Program (NCHRP) project 20-27, which developed a succession of linear referencing data models in consultation with a wide range of academicians, practitioners, and transportation policy makers (Vonderohe, Chou et al. 1997; Vonderohe, Adams et al. 1998; Koncz and Adams 2002; Koncz and Adams 2002; Koncz and Adams 2002). These efforts concentrated on identifying the most basic underlying elements in linear referencing systems and methods in order to provide a generic data model. To eliminate known difficulties stemming from differences in terminology (Dueker and Vrana 1992), these researchers comprehensively defined terms, concepts, and relationships that could apply across application areas and geographic scales of operation. Although these models comprehensively define many objects and relationships, it has been noted that "a literal interpretation of the NCHRP model would be too difficult to . . . implement" (Scarponcini 2001).

While some have concluded that a single unified linear referencing system could meet the needs of all transportation users (Fletcher, Expinoza et al. 1998), much of the NCHRP and other linear referencing modeling work has focused on the decoupling of topological, graphical, positional, and attribute characteristics of transportation objects to facilitate data sharing within enterprises (Kiel, Pollack et al. 1998; Dueker and Butler 2000) and the translation of locations between linear referencing methods (Scarponcini 2002). Another model suggests that the attributes that would traditionally be referenced to the network can be the primary object to be stored in the spatial database, while the location and shape information is encapsulated with the attribute (Sutton and Wyman 2000). Lastly, other efforts have concentrated on identifying essential data models that allow for flexible definitions of (and relationships between) transportation database objects (Curtin, Noronha et al. 2001), including those objects related to linear referencing.

Linear Referencing and GIS

The vector data model that has dominated the application of geographic information science (GIScience) since the inception of the discipline is widely recognized as an extraordinarily useful data structure for transportation systems and other network processes (Curtin 2007b). The ability to reference events to features in that data structure has long been identified as an essential functionality within GIS (Nyerges 1990). Several GIS software packages currently offer tools to assist in the generation of spatial features and events for the purpose of linear referencing (Goodman 2001). The documentation for such tools, however, is focused primarily on how to create the events and define the measures within their systems (ESRI 2001, ESRI 2003). There is very little—if any—insight into how these events and measures should be captured, analyzed, or maintained. It is difficult for users to implement linear referencing without a well-defined process to follow.

Historically, the practitioners of linear referencing in large transportation agencies knew that distance measurements collected in the field (sometimes using measuring wheels or other highly accurate distance measuring tools), and stored in tabular form (outside of the GIS), were superior to the digital data representations that—for decades—suffered from a persistent lack of positional and attribute accuracy. With the proliferation of GIS over the past decade, smaller users, including government agencies and private businesses, are capable of generating the most accurate geospatial data available for their areas. Often this is more detailed than commercially or nationally available products. These smaller users often store their data in a single location, using a single spatial data format, and they maintain that data with a small staff. This spatial representation is accepted as the highest quality representation of their network. The GIS digital data *is* the datum; the well-known street intersections or other captured points stored in the GIS *are* the anchor points for linear referencing. Given this acceptance of the data stored in the GIS, there is no need for an artificial separation of the cartographic representation and the analytic network database, which has been one of the foundations of linear referencing modeling efforts.

The use of linear referencing is a way to improve the return on the investment made in adopting geospatial technologies. When a street centerline geodatabase is being captured and maintained, building an LRS is a logical step forward that expands the number and diversity of applications that can be implemented (Noronha and Church 2002). As the paucity of literature regarding the use of linear referencing suggests, however, this valuable tool is infrequently implemented by users other than major transportation agencies. The question this research seeks to address is why do these GIS departments not implement linear referencing when the tools for doing so are readily available? The authors believe that the answer lies in the absence of a clearly defined process for implementing and using linear referencing within GIS. The following section outlines such a process.

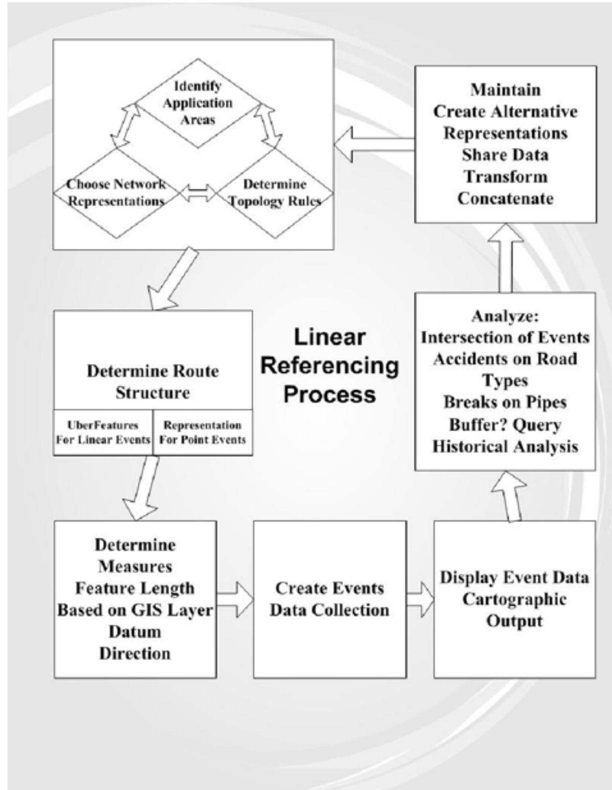


Figure 1. An iterative seven-step linear referencing process

LINEAR REFERENCING AS A PROCESS

This paper presents an iterative, seven-step linear referencing process (shown in Figure 1). The first step of this process is to identify an application to which linear referencing is pertinent and to use that information to decide what network representation should be employed and the topological rules that must be followed. The second step determines the route structure—or the underlying datum—to which events can be linearly referenced. The third step identifies the way in which measurements will be made along those routes and the fourth step defines the way in which linear events will be defined, captured, and maintained. The fifth step concerns the cartographic output of linearly referenced events, and the sixth step outlines ways of analyzing those events once they have been fully referenced. The final step in the Linear Referencing Process (LRP) is to maintain the linearly referenced data in such a way that it can be shared with other agencies, used for many different applications, and queried based on historical conditions. Each of these seven steps are outlined below and tested using a case study of the city of Richardson, Texas, a midsize city with a population of approximately 100,000 in the Dallas/Fort Worth metropolitan area.

Determine Application, Network Representation, and Topology

As described previously, myriad applications can benefit from the implementation of linear referencing. Although it would be ideal if all applications could rely on the same measurement techniques, use the same network databases, and employ the same types of network analyses, this is simply not the case. For example, road networks and municipal water networks are fundamentally different in many ways. Flow in road networks concerns independent mobile entities (cars, trucks, bicycles, etc.), while flow of water through a network of pipes is determined by demand, pressure, and elevation among other factors. Similarly, fundamental differences exist in the analytical and cartographic needs for electrical networks, gas or oil pipeline networks, or river networks. Although they all depend on a network structure, the attributes and the analytical methods associated with these different network types require different linear referencing specifications.

Therefore, the first step in a linear referencing process is to define which network datasets (and what representations of those networks) are to be employed for the application at hand. In some cases, several different representations are available for the same network, and these competing representations may differ based on their source, their coverage, their attributes, or their topological structure. Moreover, several network datasets may need to be used together, such as employing both the water distribution network and the road network for the emergency response to an incident such as a water main break.

As an example of determining application, network representation, and topology for linear referencing, we turn to the case study area—the city of Richardson—to examine a street centerline–based application for pavement management. In the city of Richardson, the road centerline was originally generated from corrected Topologically Integrated Geographic Encoding and Referencing (TIGER) data that had been provided by the North Central Texas Council of Governments (NCTCOG). It has since been modified significantly; most recent changes involve the use of six-inch orthophotos, a practice becoming more common among municipalities as high-resolution imagery is becoming increasingly available. As such, the GIS road network representation is generally trusted to be the most accurate positional reference data within the municipality. In terms of linear referencing, this means that the road network itself can serve as the datum, unlike in most theoretical linear referencing data models that require a separate datum to compensate for the inaccuracy of the centerline dataset (see Figure 2).

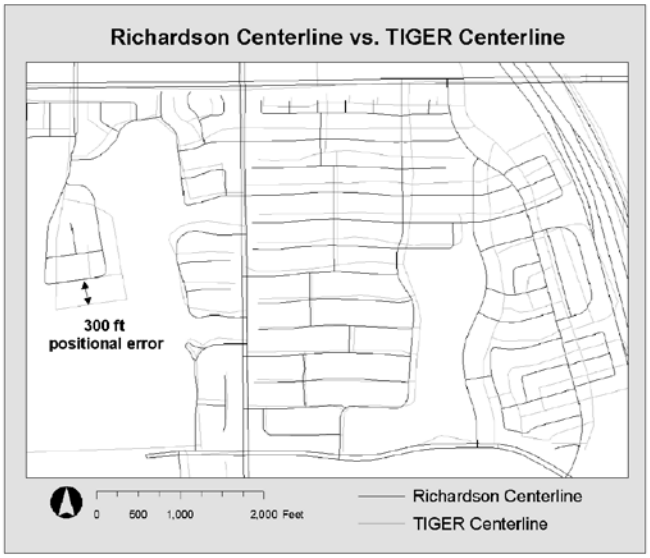


Figure 2. TIGER streets and corrected Richardson streets

One common application for the street centerline database is the inspection and maintenance of the pavement on those streets. In Richardson, the street department is provided with a cartographic representation of the street centerline file, which it uses to identify pavement surface types (concrete, asphalt, or concrete with asphalt overlay). These attributes are then input into the GIS as linearly referenced events for the purpose of querying the street centerline database to determine the amount and age of each pavement type, and subsequently to support street maintenance projects.

Given that an application has been chosen—in this case, pavement management—the specific street centerlines representation must be chosen. There can be significant differences among representations of the same road network, even within the same municipality or agency. In this paper, we discuss two primary network representations: a simple centerline representation and a quasi-carriageway representation. A simple centerline representation reflects the most common road network structure: a fully planar TIGER-based representation with centerline splits at all feature intersections, regardless of the nature of the actual street intersection at that point. Given the history of this network structure, it is the most commonly found across applications and agencies, although nonplanar alternatives exist, and several researchers have recognized the increasing demand for GIS-T data models that support lane-level operations rather than forcing all applications to conform to centerline or carriageway representations (Fohl, Curtin et al. 1996; Miller and Shaw 2001). The single centerline representation is generally used for generating road length measurements and for cartographic applications. In contrast, a quasi-carriageway centerline representation is a more detailed version of the network with individual features representing the flow of travel along major roads, particularly those that are divided by central medians (as shown in Figure 3).

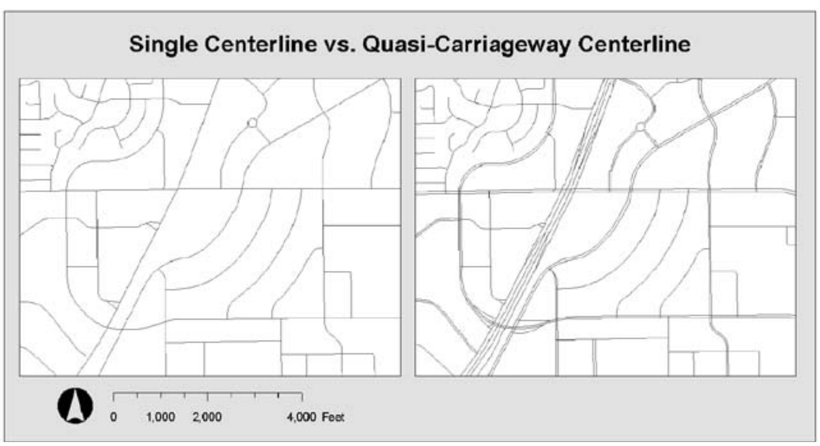


Figure 3. Simple centerline vs. quasi-carriageway

Generally, medians only separate major city roads such as arterials and collectors, therefore most residential streets have a single centerline representation. In some instances, however, roads may alternate from a single to a double representation because of the lack of or presence of a median (thus the "quasi" for not all streets are represented

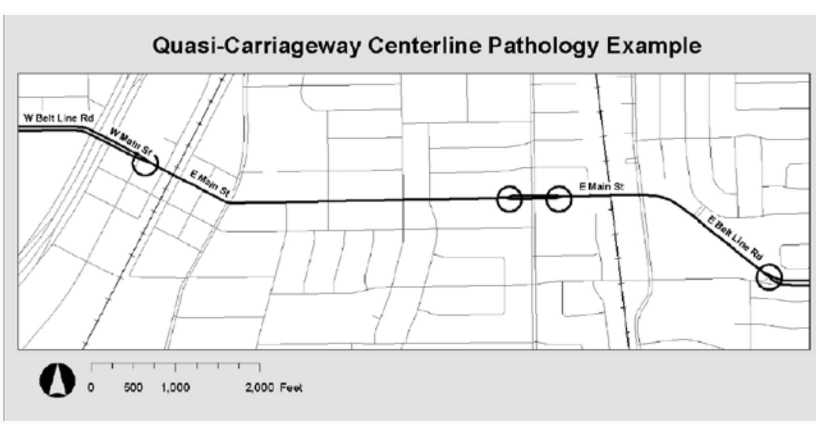


Figure 4. Quasi-carriageway pathology example

with two features showing the different directions of travel). Such network pathologies can cause difficulties in the determination of what constitutes a route in the linear referencing system (see Figure 4).

Although Richardson does not have any one-way streets, the carriageway or quasi-carriageway representation may be more appropriate for areas with large numbers of one-way streets to maintain information about traffic-flow direction. However, Richardson also maintains traffic-flow direction attributes, represented by bidirectional flow in single line segments and flow as going either with or against the digitized direction where double lines are used for street segments. Additionally, street attributes are appended directly to the centerline segments (such as block ranges for address geocoding, speed limits, pavement types, and rights-of-way, among others).

Officials in Richardson determined that it was imperative that the database not be segmented any more than necessary to reduce the network database size and to encourage accurate routing across the network. Therefore the topology of the quasi-carriageway representation had to differ from the fully planar representation, so that no segmentation of features should occur unless there is a physical intersection of streets where traffic can flow from one street to another. In cases where there are bridges, tunnels, overpasses, or other split-grade intersections, there should be no split of the features.

To summarize, for the case study of Richardson, the first step in implementing linear referencing was to choose an application (pavement management), to determine the best network representation (quasi-carriageway), and to determine the appropriate topology (nonplanar at grade-separated intersections).

Determining Route Structure

The next step—and perhaps the most challenging—in the process of linear referencing is the determination of the route structure. In this research we define a route as the largest individual feature that can be uniquely identified and to which events can be linearly referenced. This definition differs from the common notion of a route such as a bus route that may traverse several or many different features along an established course of travel. Additionally, this definition of route is mirrored most closely by the NCHRP 20-27 definition of an "Anchor Section." Features such as roads, railroads, creeks, and, ultimately, any linear feature can become the underlying element of a route.

Although routes could be created from individual street segments in the network database, this would eliminate one of the primary benefits of linear referencing, that an event spanning many street segments can be maintained as a single object in the database. To avoid this, the composite set of road segments that will constitute a route should be longer than the events to be referenced. For example, the pavement type of a street may often be the same for the full extent of a road within the city limits. Ideally, all the segments that make up that road should constitute a single route, so that one event can define the pavement type for that route. Thus, the determination of what constitutes a route must be made with regard to the largest event along that route to minimize unnecessary event segmentation.

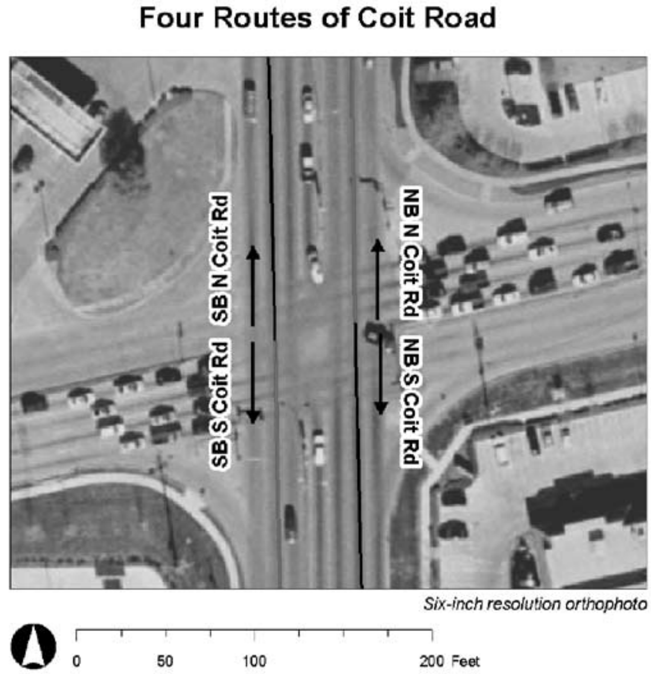


Figure 5. Route structure example

The use of an entire road spanning the geographic study area is the approach that is widely used in statewide linear referencing systems (Texas Department of Transportation 2003). However, when a single road feature has two or more names associated with it, multiple routes could be established based on the street names. Similarly, if applications (such as address geocoding) will depend on directional prefixes on the roads, it may be desirable to have routes that correspond to both street name and direction. Moreover, if a carriageway or quasi-carriageway network representation had been chosen in step one of the linear referencing process, then each of the carriageways may be a separate route, even if the street name and direction prefix are the same.

In the Richardson case study, the underlying street network database has a quasi-carriageway representation. Topologically, splits on this database are enforced only where at-grade intersections occur, where road names change, or where address block numbers change.

Given these representational choices, the routes for linear referencing have been based on the street name with directional identifiers. When carriageways (double lines) exist, a separate route is established for each direction of travel (*e.g.,* northbound and southbound). Because residential roads are all single lines, each full street name/direction prefix constitutes one route. When a street has no directional prefix, the entire street becomes a route. For example, the two sets of segments that make up Coit Road were actually split into four separate routes according to the following route-name attribute: NB N Coit Rd, SB N Coit Rd, NB S Coit Rd, and SB S Coit Rd (see Figure 5).

In summary, the route structure determination decision depends on all the decisions made in the prior step of the linear referencing process in addition to decisions regarding what attributes are to be linearly referenced. This necessitates an evaluation of the trade-off between the level of detail that one wants to reference against and the concomitant increase in the number and segmentation of routes.

Determining Measures

Once the route structure is determined, the third step in the linear referencing process is to determine measures along those routes. There are three primary considerations when setting measures along routes: (1) the unit of measure that is most appropriate, (2) the source for the measure values, and (3) the direction of increasing measure values. The most appropriate unit for measures along routes is a function of the application for which the linearly referenced features will be employed. Because unit conversions are a commonplace function in GIS software, changes in units require only that both the measures along the routes and the measure information associated with the event data remain consistent. However, if more than one unit is required for the same routes, this may require that a second set of routes be maintained. For some applications, the measure along a feature may be given as a percentage of the distance along the feature rather than as an absolute distance.

As discussed previously, the source data for measure values is a subject of substantial debate. Historically, data collected in the field and stored in databases external to the GIS were considered substantially higher in quality in terms of spatial accuracy than the digital spatial data employed by the GIS. This fact led linear referencing researchers to develop well-defined objects (such as anchor points and anchor sections in the NCHRP 20-27 model) that were based on the well-defined location of points and segments collected in the field. Measures were computed along the network based on their distances from these objects. These measures could be associated with cartographic representations of the network, but differed from the distance values computed internally by the GIS. Today, the digital data landscape has completely changed in this respect. The data maintained in GIS departments is almost universally considered the highest-quality spatial data available, and any improvements or corrections in feature locations are almost immediately transferred to the GIS database. Additionally, consortia of municipalities and nongovernmental organizations collaborate extensively on data collection and maintenance. Therefore, those implementing linear referencing today are free to base their measures on the positions and lengths of features as they are computed and stored within the GIS. There is no need to maintain a separate set of features from which measures are computed. Once again, the GIS data *is* the datum.

The third consideration when determining measure values is the direction of increase of those values. Once again this depends on the previous steps in the linear referencing process. Generally speaking, the direction of increase should be consistent with the needs of the application chosen in the first step of the process, and it should be logically consistent with both the topological network design and the route structure determined in step two of the process. For example, consider an application concerned with a river network and multiple routes associated with different branches of the network. Because activities along the river network are likely to be associated with the flow of the river, the measures along the routes should be consistently associated with either the upstream or the downstream direction of the river network. Similarly, if routes are designed to conform to the direction of travel along a street network, then increasing measures may most appropriately conform to those directions. However, if applications associated with addresses are of primary interest, then the direction of increasing measures should likely conform to increasing address ranges.

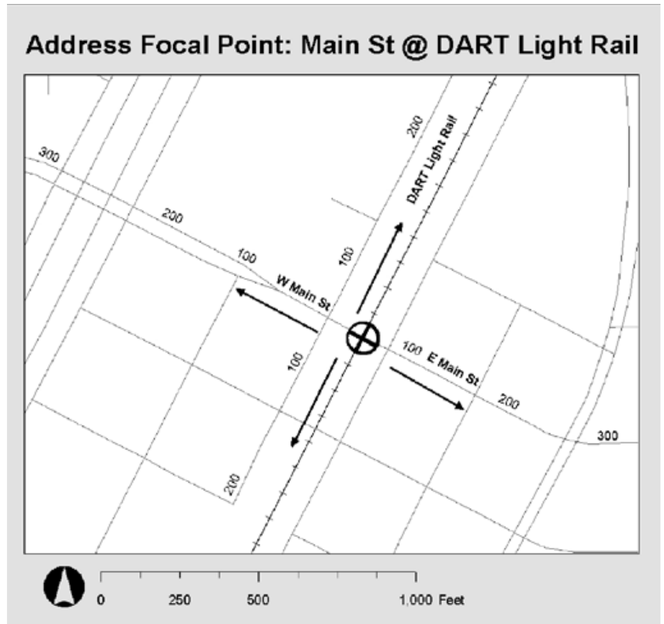


Figure 6. Measures increasing from the focal point

Turning to the case study area, in the city of Richardson, measure values were based on feature length because the GIS data is accepted as the datum and the geometric lengths of the segments are accepted as sufficiently accurate for most city needs. In Richardson, route measure units were based on the projection of the data (state plane) that customarily employs feet as the unit of measure. For Richardson, route measures were designed to increase in accordance with the addressing scheme for the city. More specifically, measures increased from the origin point of the city's street grid that is the intersection of Main Street and the Dallas Area Rapid Transit (DART) rail station located in the historic downtown section (shown in Figure 6). This is the same point used as the dividing point for routes.

Create Events

When the first three steps of the linear referencing process have been completed, a set of routes have been built from the underlying network for the chosen application. These routes are informed with measure information based on application needs and topological structure. The next step in the process is the collection of

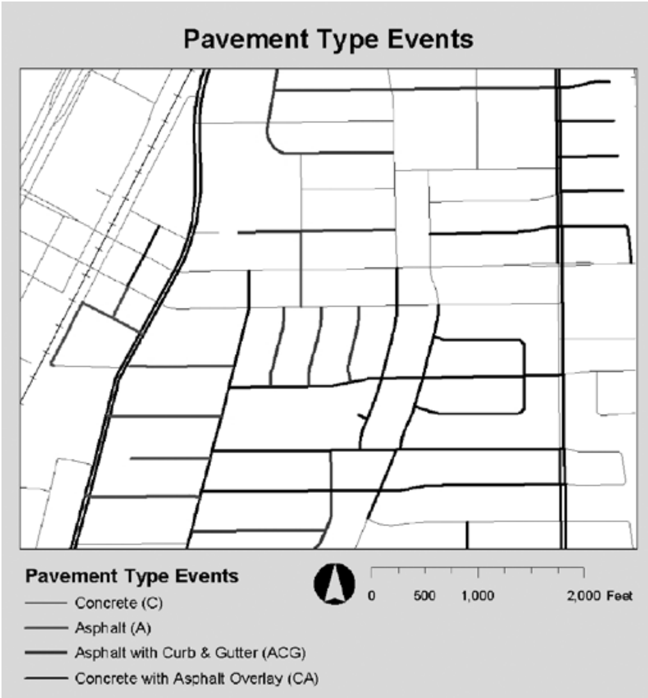


Figure 7. Pavement types represented as linear events

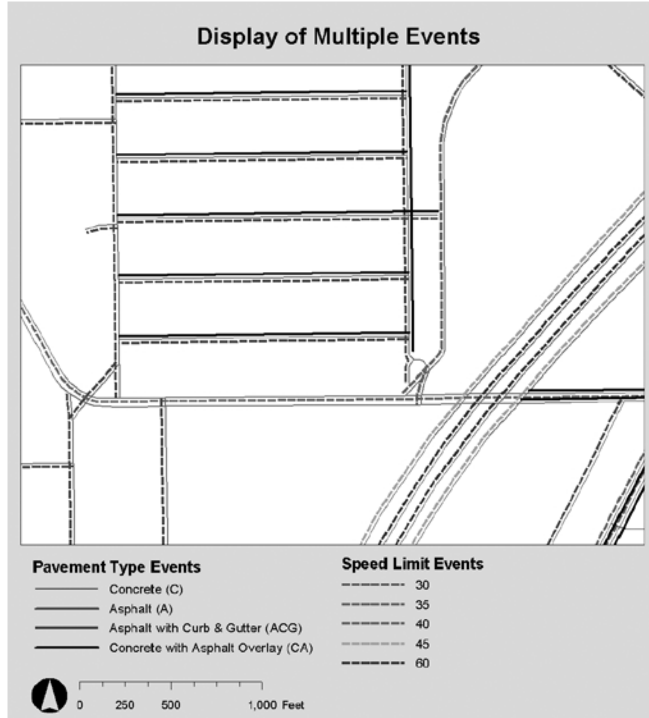


Figure 8. Cartographic display of related events

event data to be referenced to the newly created routes. For many users, this step is what first comes to mind when considering linear referencing. However, without having completed the first three steps of the process as outlined previously, a great deal of effort in event data collection could be wasted.

Event data are occurrences along the network. Events can be point or linear in character. Point events represent some object at a specific measure along a linear feature. Examples include traffic accident locations or traffic control devices such as signs or signals. Linear events often correspond to objects that have a consistent attribute along the network. Examples include pavement type or condition, speed limits, traffic volumes, or pipe widths along a water network. An event is known as a "traversal" in the NCHRP 20-27 efforts.

Event data can be collected or created in many different ways. Events can be digitized from a range of cartographic products including both paper maps and aerial or satellite photographs. Events can be collected in the field either by direct observation of attributes or locations by personnel, or with GPS receivers that capture locations for subsequent input to the GIS as events. Custom software tools exist to facilitate event data collection, conversion, and maintenance. Existing point data can also be converted to point events, or event tables can be populated manually based on known locations and attributes. As needed, event data can be exported as a set of stand-alone features; otherwise, it can remain in tabular format. Perhaps most important, whichever method of event data collection is chosen, the events must be structured in such a way that they can logically be associated with the routes determined in the second step of the process.

In the case of Richardson, the initial objective for implementing linear referencing was to more accurately represent surface pavement type. Data on the locations of different pavement types had been collected in the field and recorded on a paper map. This data had been transferred to the network database, but was represented inaccurately since road segments had not been split for this purpose. Therefore, the pavement type attributes had to be approximated to the nearest intersection. With the implementation of linear referencing, more precise pavement type events could be associated with the previously determined routes without further splitting the network features solely for the purpose of attribute differentiation (shown in Figure 7).

Display Event Data; Cartographic Output

One of the most powerful arguments for the implementation of linear referencing within a municipality is the ability to accurately display the event information for in-the-field use by employees, or for higher level analysis and decision making. The ability to display multiple attributes associated with networks provides both opportunities and challenges. The increased information available for display can provide new insight into the problems under examination, but can also lead to poor cartography because of graphical clutter and information overload. Therefore, the next step in the process of linear referencing is to carefully choose the parameters for display of the linearly referenced information.

The decisions regarding display of event data depend on several factors, including the media on which the data will be displayed, the scale (or scales) at which the data will be displayed, and the computation of event locations within the GIS. Their

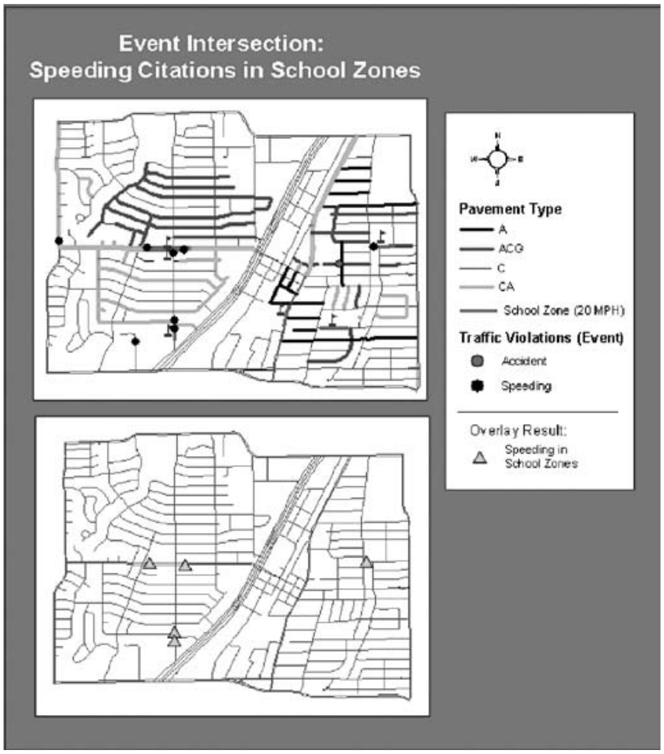


Figure 9. Intersection of linear and point events

subsequent display is a process that is often referred to as dynamic segmentation. One visual benefit of dynamically displaying event data is the ability to display multiple linear events along the same feature, using varying event offsets. Common examples of this practice include subway maps and bus route maps. Richardson used the pavement type events and speed limit events to create a single cartographic display of the locations of these two attributes more clearly than could have been accomplished using other cartographic procedures (see Figure 8).

Analysis with Linear Referencing

The ultimate goal of implementing linear referencing (or any other process) in a GIS is to increase the ability to perform a diverse set of analyses. With routes and event data in hand, analysis can then be performed on the event data, through techniques such as overlays, intersections, and other techniques that are part of the linear referencing capabilities of most GIS software. For example, in Richardson, the association between traffic violations and school zones was explored through an intersection of speed limit linear events (specifically those events associated with school zones) and traffic violation point events (as shown in Figure 9).

Linear referencing allows an entirely new set of database queries to be made that differs from queries based on the underlying network. For instance, the storage of data as event tables enables historical queries if events are date-stamped. However, while significant analytic capability is added through the linear referencing process, other traditional GIS analytic capabilities are lost. One example is the loss of traditional road network functions, such as shortest path determination or routing, because of the loss of nonplanar topology that results when the segments that comprise the routes must be merged to satisfy the need of a large-enough route to minimize event differentiation.

Table 1 contains a comprehensive list of common GIS analytical tools and describes whether or not these tools can be used with linearly referenced event data. It is important to note that the tools for network analysis, geostatistical analysis, and geocoding currently cannot be used with the linearly referenced events. These are functions that are universally used for network-based applications, and their extension for use with linearly referenced data would represent a substantial advance for network analysis.

Data Maintenance

To keep the newly created linear referencing system functional, it is important that the route and event data be maintained properly. As changes are made to the original road file, the same changes must also be reflected in the routes. There are extant tools for setting appropriate topology rules, which, in turn, enable multiple route feature classes using different linear referencing methods to be built and maintained on a common reference layer of roads. If the original roads and all associated routes are united in a topology, this also enables the simultaneous editing of these multiple feature classes.

Furthermore, measure values need to be maintained if roads are ever removed or have their course altered over time. Or if even more precise measure data becomes available, routes can be calibrated to reflect this new data. While the maintenance of the linearly referenced data and the underlying network may not be the most fascinating of tasks, it is necessary to keep an implementation of linear referencing functioning.

CONCLUSIONS

Although there are a multiplicity of high-quality theoretical models and methods of linear referencing, and a substantial number of tools for implementing linear referencing in a GIS context, there has until now been no explicit process for implementing linear referencing. This paper presents a comprehensive process for linear referencing that consists of outlining the application for which linear referencing is intended, defining the nature of the underlying network, identifying the underlying routes, specifying a system to measure locations along those routes, collecting and storing event data, performing analysis with those events, and maintaining the linear referencing system. This seven-step linear referencing process is intended as the basis from which any application of linear referencing can proceed. It is hoped that this structure will allow the extraordinarily useful set of linear referencing tools to be more widely accepted among GIS users, and will encourage GIScientists to more closely examine the processes behind these tools and thus increase the ability to perform robust geographic analyses.

Table 1. Event Analysis with Different Tools

Tools	Event Analysis	Tools	Event Analysis
1. Analysis Tools		6. Data Management Tools	
➢ Extract		➢ Database	No
o Clip	Yes	➢ Domain	No
o Select	Yes	➢ Feature Class	
o Table Select	Yes	o Calculate Default Cluster Tolerance	Yes
➢ Overlay			
o Intersect	Yes	o Calculate Default Spatial Grid Index	Yes
o Union	No		
➢ Proximity		o Integrate	Yes
o Buffer	Yes	➢ Feature	
➢ Statistics		o Add XY coordinate	No
o Summary statistics	No	o Check Geometry	Yes
2. Conversion Tools		➢ General	
➢ From Raster	No	o Merge	Yes
➢ To dBASE	No	o Copy	Yes
➢ Geodatabase		o Append	No
o Feature Class to Geodatabase	Yes	➢ Fields	
o Table to Geodatabase	Yes	o Add Field	Yes
➢ To Raster	No	o Calculate Field	Yes
3. Spatial Statistics Tools		➢ Generalization	
➢ Analyzing Pattern		o Dissolve	No
o Average Nearest Neighborhood	Yes	➢ Indexes	
o High-Low Clustering	Yes	o Add Attribute Index	Yes
o Spatial Autocorrelation	Yes	➢ Join	
➢ Mapping Custers		o Add Join	Yes
o Cluster and Outlier Analysis	Yes	➢ Layers and Tables View	
o Hot-Spot Analysis	Yes	o Make Feature Layer	Yes
➢ Measuring Geographic Distribution		o Make Query Table	Yes
o Central Feature	No	o Make XY Event Layer	Yes
o Directional Distribution	Yes	o Select Layer by Location	Yes
o Linear Directional Mean	No	o Select Layer by Attribute	Yes
o Mean Center	Yes	7. Cartography Tools	
o Standard Distance	Yes	➢ Masking Tools	
➢ Utilities		o Cul-De-Sac Masks	No
o Calculate Area	No	o Feature Outline Masks	Yes
o Collect Events	Yes	o Intersecting Layer Masks	Yes
o Count Rendering	No		
o Export Feature Attribute to ASCII	No	8. Network Analyst Tools	No
o Z-score Rendering	Yes	9. 3-D Analyst Tool	No
4. Geocoding Tools	No	10. Geostatistical Analyst Tools	No
5. Data Interoperability Tools	No	11. Spatial Analyst Tools	No

About the Authors

Kevin M. Curtin is Associate Professor of Geography at George Mason University. His teaching and research interests include Network GIS, Transportation Geography, and Optimal Facilities Location Science.

Corresponding Address:
Department of Geography
George Mason University
4400 University Drive MS1E2
Fairfax, VA 22030
curtin@gmu.edu

Greta Nicoara has a master's degree in Geospatial Information Sciences from the University of Texas at Dallas. She is currently a Senior GIS Analyst/Programmer with the City of Richardson, Texas.

Rumana Reaz Arifin is a Masters Student in Grographic Information Science at the University of Texas of Dallas.

Corresponding Address:
Program in Geographic Information Sciences
School of Economic, Political and Policy Sciences (GR 31)
University of Texas at Dallas
800 West Campbell Road
Richardson, TX 75080-3021

References

Curtin, K. M. 2007a. Network data structures. Encyclopedia of geographic information science. K. Kemp, ed. Thousand Oaks: Sage Publications.

Curtin, K. M. 2007b. Network modeling. The encyclopedia of geoinformatics. H.

Karimi, ed. Hershey: Idea Group Publishing.

Curtin, K. M., V. Noronha, et al. 2001. ArcGIS transportation data model. Redlands, CA: Environmental Systems Research Institute.

Dueker, K. J., and J. A. Butler. 2000. A geographic information system framework for transportation data sharing. Transportation Research Part C-Emerging Technologies 8(1-6): 13-36.

Dueker, K. J., and R. Vrana. 1992. Dynamic segmentation revisited: a milepoint linear data model. Journal of the Urban and Regional Information Systems Association 4(2): 94-105.

ESRI. 2001. Linear referencing and dynamic segmentation in ArcGIS 8.1. Redlands, CA: 56

ESRI. 2003. Linear referencing in ArcGIS: practical considerations for the development of an enterprise-wide GIS. Redlands, CA.

Federal Highway Administration. 2001. Implementation of GIS based highway safety analysis: bridging the gap. McLean, VA: U.S. Department of Transportation, 104.

Federal Transit Administration. 2003. Best practices for using geographic data in transit: a location referencing guidebook. Washington, D.C.: U.S. Department of Transportation, 211.

Fletcher, D., J. Expinoza, et al. 1998. The case for a unified linear reference system. Journal of the Urban and Regional Information Systems Association 10(1).

Fohl, P., K. M. Curtin, et al. 1996. A Non-planar, Lane-based, Navigable Data Model for Intelligent Transportation Systems. International Symposium on Spatial Data Handling, Delft, The Netherlands, International Geographical Union.

Goodman, J. E. 2001, November 28, 2001. Maps in the fast lane—linear referencing and dynamic segmentation. April 23, 2004, http://www.directionsmag.com/article.php?article_id=126.

Kiel, D., J. Pollack, et al. 1998. Issues in adapting linear referencing systems for transportation applications: current practice and future outlook. Journal of Computing in Civil Engineering 12(2): 60-61.

Koncz, N. A., and T. M. Adams. 2002. A data model for multi-dimensional transportation applications. International Journal of Geographical Information Science 16(6): 551-69.

Koncz, N. A., and T. M. Adams. 2002. A data model for multi-dimensional transportation location referencing systems. Journal of the Urban and Regional Information Systems Association 14(2): 27-41.

Koncz, N. A., and T. M. Adams. 2002. Temporal data constructs for multidimensional transportation geographic information system applications. Transportation Research Record 1804: 196-204.

Miller, H. J., and S. Shaw. 2001. Geographic information systems for transportation. New York: Oxford University Press.

Noronha, V., and R. L. Church. 2002. Linear referencing and other forms of location expression for transportation. Santa Barbara: Vehicle Intelligence & Transportation Analysis Laboratory, University of California, 26.

Nyerges, T. L. 1990. Locational referencing and highway segmentation in a geographic information system. ITE Journal, March: 27-31.

Scarponcini, P. 2001. Linear reference system for life-cycle integration. Journal of Computing in Civil Engineering 15(1): 81-88.

Scarponcini, P. 2002. Generalized model for linear referencing in transportation. Geoinformatica 6(1): 35-55.

Sutton, J. C., and M. M. Wyman. 2000. Dynamic location: an iconic model to synchronize temporal and spatial transportation data. Transportation Research Part C-Emerging Technologies 8(1-6): 37-52.

Texas Department of Transportation. 2003. Texas reference marker (TRM) system user's manual, 493.

Vonderohe, A., T. Adams, et al. 1998. Development of system and application architectures for geographic information systems in transportation. Washington, D.C.: National Cooperative Highway Research Program, Transportation Research Board, 23.

Vonderohe, A., C. Chou, et al. 1997. A generic data model for linear referencing systems. Washington D.C.: National Cooperative Highway Research Program, Transportation Research Board.

When Data Sharing Becomes Institutionalized: Best Practices in Local Government Geographic Information Relationships

David L. Tulloch and Francis Harvey

Abstract: *Through a series of case studies examining local government data sharing practices, the authors compiled a list of best practices. These best practices are described and analyzed to show keys and contradictions in data sharing practices, highlighting long-term opportunities and problems within data sharing communities. Notably, the paper examines a ripple effect that results from the initial decision to constrain data uses through a policy that charges for data. The authors conclude pointing out the crucial need for continued education and coordinated efforts to advance emerging local government SDIs.*

INTRODUCTION

As agencies function within the context of an increasingly digitally interconnected world, the ability for information to flow from a source to many other unanticipated individuals has been accelerated, especially through networks of other interested individuals (Watts 2003). The response to this phenomenon of many people involved with local government geographic information systems (GIS) continues to be apprehension and resistance. The resulting data-sharing networks exhibit a variety of patterns that reflect contingencies and exigencies in a particular local government.

Whether as a means of data dissemination or acquisition, data sharing has become an essential element of local government GIS activities. Despite the prominence of this activity and its centrality to the day-to-day function of many local government systems, decisions about these activities are rarely made following a thorough consideration of the organizational and political complexity. This paper examines the techniques used in a variety of communities and extracts from those a summary of best practices for data sharing in local government. These best practices are offered as guidelines borne out by the outcomes experienced under a variety of circumstances, but as any best practices they remain admittedly suggestive more than absolute.

The importance of existing political, institutional, professional, and legislative relationships is emphasized in this paper. These relationships largely determine the geographic information activities at local government agencies. Sharing and coordination are to a large extent informal activities that correspond to these relationships. Formal arrangements are made only at the final stages of establishing sharing or cooperation agreements; basically, they manifest themselves at the conclusion to satisfy internal procedures and/or legal requirements. This understanding poses an interesting conundrum for developing local government participation in the National Spatial Data Infrastructure (NSDI).

Given the importance of informal agreements, the primary best practices question arises: How do local governments successfully share geospatial data and coordinate geographic information activities? Improved understanding of the issues, potentials, and problems that local governments face is key to establishing programs and policy support to address local government needs and to develop better opportunities for them to participate in geographic information infrastructures connected to any spatial data infrastructure (SDI).

BACKGROUND

Data sharing is much more than a simple technique for accessing geographic informaton (GI); it is a unique form of institutional interrelationship that is necessary for promoting many of the most important benefits of public GIS applications (Tulloch and Epstein 2002, General Accounting Office 2003). Perhaps the grandest potential public GIS application would be an SDI, which can only be accomplished with significant advances in data-sharing practices and a fundamental improvement in the understanding of the ways in which these practices are employed.

Many of the larger GIS initiatives (both formal and informal) are based on the concept of many data producers working collectively to develop large volumes of data combined to describe relatively large areas with an expectation that those data will be widely available to potential users. Included among these initiatives are a National Spatial Data Infrastructure (NSDI), a Global Spatial Data Infrastructure (GSDI), INSPIRE, and the Digital Earth initiative. Each of these would require a large volume of existing data and substantial accessibility to the data.

The nature of data access in the United States can be quite complex, with a variety of state and federal laws governing public organizations, and constantly changing markets governing commercial distribution (Azad and Wiggins 1995). Initially, access to data was allowed largely under ad hoc arrangements, but as GIS institutions continue to grow, more organizations are developing policies and practices that formalize the ways in which their geospatial data can be used and disseminated. The different forms of access are now generically called *data sharing*.

Spatial data sharing is defined as the (normally) electronic transfer of spatial data/information between two or more organizational units where there is independence between the holder of the data and the prospective user. The transfer may be in the form of periodic bulk transfers, routine daily transfers, or on-line access driven by individual transactions. The participants may be separate organizations or may be departments within the same organization. For our purposes, the distinguishing characteristic of spatial data sharing is that there be an arm's-length exchange or transfer. (Calkins and Weatherbe 1995, 66)

The arguments for data sharing are many, but a traditional understanding of its usefulness is often based on the fact that sharing allows data, otherwise isolated and underutilized, to be used repeatedly for many purposes, thus increasing their value without increasing their cost (MacKaay 1982).

The value and social utility of geographic information comes from its use. Sharing of geographic information is important because the more it is shared, the more it is used, and the greater becomes society's ability to evaluate and address the wide range of pressing problems to which such information may be applied. (Onsrud and Rushton 1995, xiv)

Despite the apparent values of data sharing, many barriers to sharing still exist, from Craig's (1995) broadly presented "institutional inertia" to Kevany's (1995) detailed list of 30 geographic information factors. More recent, Nedovic-Budic, Pinto, and Warnecke (2004) have examined the mechanisms and means through which data sharing currently occurs with a particular eye to the motivations behind sharing. An added complexity is the unique nature of human behavior that When de Montalvo (2003) seeks to take into account as an explanation of data-sharing activities and an understanding that should further shape policies.

The identification of best practices has become a common approach in business and public administration. Many have traced best practices to the popular press book, *In Search of Excellence* (Peters and Waterman 1982). The approach, however, has grown into a tool for studying a variety of situations and identifying successful ways of addressing those situations. Best-practices reporting has become a particularly useful tool for capturing options available to managers or decision makers, and identifying appropriate applications for these practices.

METHODOLOGY

Because much of the methodology of this paper has been previously described (Harvey and Tulloch 2003, 2004, 2006), we will provide only an abbreviated version for background. The project used two primary methods of data collection:

a) Interview-based local government data-sharing case studies and
b) Multimodal local government data-sharing surveys.

The methodology is divided into three distinct phases. In phase one, we examined the experiences of data-sharing participants with local sharing and coordination arrangements. Objectives of this phase were to determine strategies and key issues in demonstration projects that have worked and find out what strategies failed and the problems that were resolved and continue to have impacts.

In phase two, we conducted interviews at a variety of local governments across the United States to determine how they share geospatial data and coordinate geographic information activities. These interviews were held with numerous local government agencies. Interviews throughout the entire data-sharing network provided substantive insights into the contexts and practices of sharing and coordination in situ. The local governments whose GI sharing practices were studied were selected with great care. We included large and small local governments that represented a diverse geography, and we worked to ensure that legislative and political differences at the regional and state levels could be identified for comparison.

The in-depth interview case studies were held in six areas around the United States. These in-depth interviews of multiple participants within a single data-sharing network provided great detail that was complemented by the more tightly focused polling of local governments in the surveys. For each series of case-study interviews, the authors identified a significant spatial data-sharing network that relied heavily on local government data. Several of the case studies were chosen based on their use as Federal Geographic Data Committee (FGDC) demonstration projects. All the examples focused on local government data that was shared in a variety of ways with different users or organizations. The interview processes generally began with an extended interview with one of the networks most central GIS coordinators and then moved on to interviews with other data providers and data users within that network.

FINDINGS

This section identifies specific repeated lessons from the surveys, interviews, and case studies. We identify those outcomes most broadly relevant to local government GIS data practices. The findings described in this section are derived directly from the data-collection processes described in the methodology. Some are based on experiences of many different organizations, while others are derived from the experiences of the most successful data sharers.

These findings reveal some recurring keys to success in communities where data sharing has thrived, while also illustrating some complications that could confuse a causal observer. Based on these findings and the related results of this research project (Harvey and Tulloch 2003, 2006), the authors have developed a list of data-sharing keys and contradictions that reflect data-sharing best practices.

As is frequently pointed out by critics of "best practices" as a form of knowledge management (e.g., Patton 2001), no one set of best practices can exist for a complex, context-dependent, network-based activity such as data sharing. But this list provides a quick summary of some better practices and important contradictions that emerge from the project described in this paper.

SEVEN KEYS TO DATA SHARING

K1. Context matters.

Different institutions require different responses. Decisions about whether data sharing should be formal or informal may depend on where you are. No matter how badly you want to give away the data, if you work in "Tammany Hall," you can't. Some isolated systems may be able to rely on a hub-and-spoke model, while some complex multijurisdictional landscapes may require fairly sophisticated models to ensure that data is available in appropriate and fulfilling ways.

K2. Attitudes vary.

When asked whether they charged, some people asked, "Why would we?" When asked if they shared data, some people asked, "Why would we?" Some people saw the data as the source of their power. Some people saw giving it away as their source of power. Sharing scares some people. Some people have concerns about the risks involved—liability is a big concern. (But some are suggesting that liability goes up when you charge.)

K3. Charging for data can cost more than you think.

(Be wary of the ripple effect.) Sometimes when you think you are making money from your data, you are really costing more than you think. Charging for data can have primary, secondary, and tertiary economic effects that you should be aware of before you chose a restrictive access policy.

K4. Bigger is better.

Generally speaking, the larger organizations seemed more likely to share, to have developed metadata, and to be prepared to participate in a larger spatial data infrastructure.

K5. Where there's metadata, there's data.

While not all data come with metadata, we found that folks who kept metadata almost always had lots of data and it generally seemed to be data worth getting. But many agencies admitted to using standards that were not FGDC-compliant.

K6. Sometimes it's all about who you know.

A number of institutions explicitly admitted to sharing data freely with people they know and trust, while making it difficult for others to gain access. It became clear in some places that almost everybody had studied at the same school, so that even if they weren't classmates they shared the same favorite faculty. These personal connections seemed to really overcome some other limitations. It can also be about how you treat them. When you are working in a hub-and-spoke environment, you clearly need to treat the GIS dictator nice. If he or she shuts you out, it can be very hard to get back in.

K7. Sharing is easy, not sharing is hard.

Just giving your data away can turn out to be the easier and more affordable route. Copying someone else's shrink-wrap agreement and leaving your data on the Web page can be pretty simple. Dealing with lawsuits, chasing down "illegal data launderers," and negotiating ironclad license agreements can be *very* hard and unrewarding work. Some agencies seem to spend more time and energy dealing with preventing "data theft" than they make in their cost-recovery charges.

FIVE CONTRADICTIONS OF DATA SHARING

C1. Remember, to give is divine, but knowing you might get something back later is pragmatic.

Data sharing is a good thing—as good as mom, apple pie, and the flag. While economic considerations may motivate people to hold on to data or charge for it, most people recognize that data collected with public funds is a public good. Sharing public goods freely is as important to the democratic process as voting. All the same, it is undeniably important to know that when you share data, you (and/or the department) will be recognized. Recognition is an intangible currency of public administration, but one of the highest values for an organization and local politics.

Unexpected ways of receiving are also very important. An agency might pass on new data to a contractor and receive a windfall at the next budget meeting because a council member was able to ensure the development of a new park thanks to the contractor saving money on survey costs.

C2. Give the data away, but always make sure people know where it comes from.

Obviously, giving data away makes sense—for any number of reasons. Also because no one has yet been able to demonstrate a successful cost-recovery program in the United States. You also need to be sure that people know where the data came from. If they don't know where it came from, they don't know who to contact about updates, changes, and possible uses, and suggested improvements. People also don't know who to thank and recognize.

C3. Don't charge for data, but make sure people know what it costs.

What's the value of something you get for free? Nothing, claim pundits. That may be so among people who deride the responsibility of government and wish to diminish its role. However, people who know the value of good government generally appreciate its value more. Agencies that communicate the value of data they share for free can be easily recognized. Of course, determining value is a complicated affair. Be sure to be judicious and conservative in estimates. People value a good value even more.

C4. If possible, never charge fees, but make additional services available at a fair cost.

If revenue and cost recovery are important issues, instead of fees, many agencies have been successful in providing additional services to the no-cost provision of data. This can be done alone by a department, but public-private partnerships can be important ways to create some jobs in the community and strengthen local bonds. Remember, that data is just the smallest part of any GIS. People and analysis are the largest parts and they require resources. You might even want to think about offering training support to give people from other departments and communities a solid start. Once they know what can be done, they will be back many times for more data and help.

C5. If you have data, remember, you don't own the room that people are eating in, you're only putting the "food" on the table.

The last time you went to a catered party (or set up the party yourself), do you remember the caterer decorating the room or telling the host or hostess how to greet visitors? Because this is unlikely to ever happen, remember that when you share data, you are only dealing with part of the activities. It may be the most important part, but remember who is using the data and its importance for their activities. Help with the data to make their work a success and the benefits will be bountiful.

To designate the manner in which each finding contributes to a best practice described later in the paper, the text in this section is marked with notations indicating which key or contradiction (*e.g.,* [K2] or [C4]) the finding most closely supports. The first part of the section describes different models for understanding local government data-sharing arrangements. After an evaluation of underlying issues impacting these models, the second part presents a set of best practices.

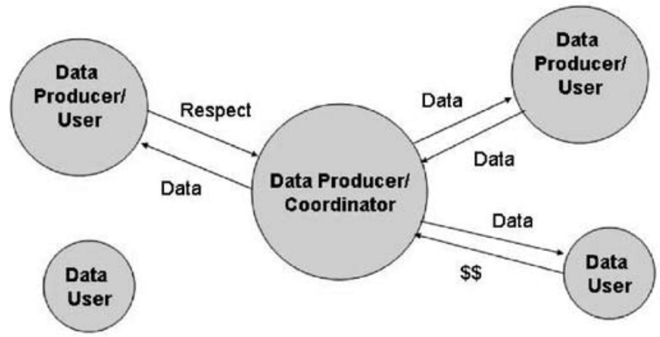


Figure 1. A conceptual diagram of the centralized network (or GIS dictator) model of data sharing. This can be a very attractive model when the central figure has strong expertise and is generous (a benevolent dictator). Less considerate or skilled coordinators can quickly turn this model into a plan for disaster.

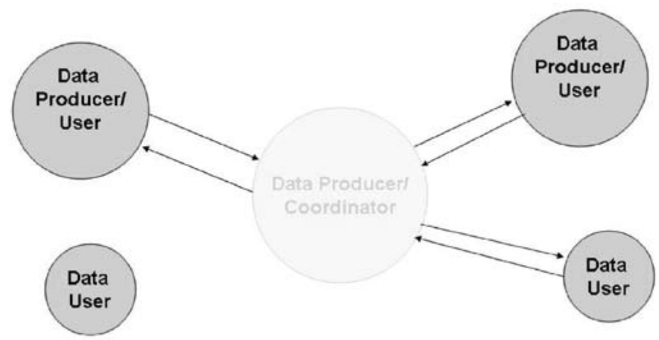


Figure 2. The sudden departure of a strong centralized coordinator (or GIS dictator) can leave other users in a difficult situation. They may even be unable to access data that was produced by a third party or decipher cryptic metadata that was considered important when the source could be consulted.

Different Structures Require Different Responses

The context within a local government for data sharing should tie into the broad issues within which an SDI, or a larger network of data users, is situated. Naturally, the operations and organization of each local government will reflect exigencies and contingencies. Certainly, for a study of best practices, this requires understanding the activities of local government agencies. We have identified some distinct data-sharing structures that form the context within which sharing often occurs (Harvey and Tulloch 2006). Understanding the contextual structure helps clarify concerns and responses described by local government agencies [K1]. For example, a centralized "hub-and-spoke" model that relies on a strong coordinator (see Figure 1) might create situations for which formal arrangements and fully compliant metadata are less important. But unexpected changes in the model, particularly the sudden departure of the central figure, can create institutional chaos.

We can contrast the centralized model with that of either the federation-by-accord or the federation-by-mandate (Harvey and Tulloch 2006). These models for federated data-sharing arrangements among local governments (as shown in Figure 2) reflect either the negotiated agreements that led to the formalization of data-sharing arrangements or were stipulated by a government agency. Federated data-sharing arrangements are the most common; not all are thoroughly formalized, but they are in a state of flux corresponding to changes in institutional, personal, and political arrangements (see Figure 3). While the inherent slack in these arrangements may seem to be a negative point from a policy point of view, practically this flux makes these arrangements enormously flexible and resilient to changes. The few stipulated arrangements exhibit both negative and positive characteristics, depending on the degree of funding for the activities and their role as a "seed" for additional data-sharing arrangements.

The Hidden Costs of Cost Recovery: The Ripple Effect (or the Cost-Recovery Cascade)

An important issue in all data sharing is financing and conflicts related to cost-recovery strategies. Economists point out that sharing information is different than sharing manufactured goods

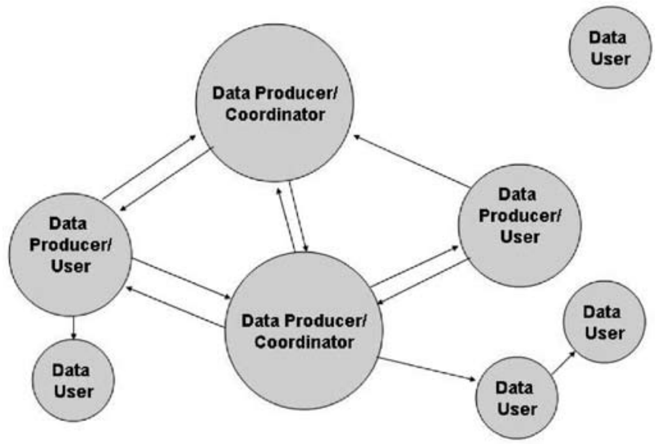


Figure 3. A conceptual diagram of the federated network model of data sharing. As it grows more complex, the federated network can create many redundancies that become very important in difficult times.

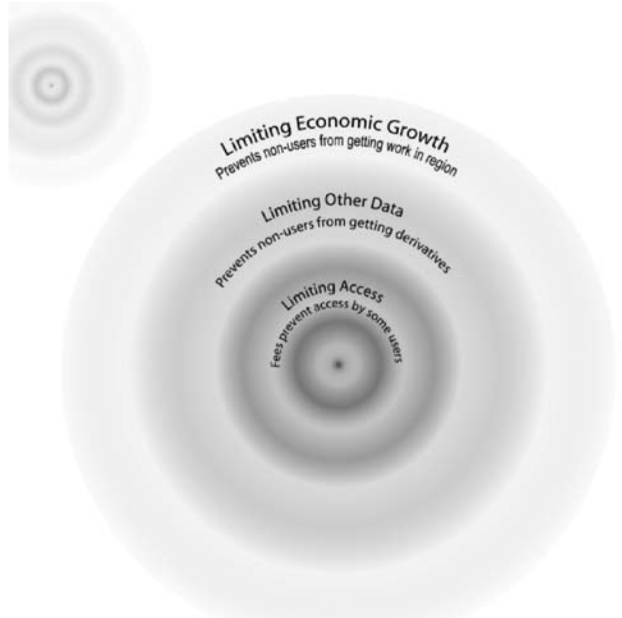


Figure 4. The ripple effect of limited access illustrates how a simple decision to limit access through the use of user fees can have far-reaching effects.

(Mackaay 1982); still, many are concerned about the costs of sharing geospatial data. Much has been written about the nature of the debate within the GIS literature regarding the appropriateness (or inappropriateness) of policies promoting cost-recovery techniques that effectively charge a fee for access to public data (Epstein 1991, Onsrud and Rushton 1995) as well as the impact of sharing data across organizations (Azad and Wiggins 1995).

This research isolated a pattern for data-sharing arrangements involving prohibitory cost-recovery policies, not previously described in the literature, caused by a cost-recovery policy. This pattern, which we'll call the ripple effect, was found, for example, in a community where one of the most important datasets (parcels) was protected by a thoroughly developed and rigorously enforced policy that charged a significant fee for access and limited redistribution. The policy included specific language that restricted the ways that these data could be incorporated into other forms of analysis that might eventually be released publicly. As it played out, this policy had three different levels of impacts, many of which were not immediately obvious, but all of which demonstrate the ways that such a policy stifles activities throughout a GIS community [K3].

The primary (and most discussed) impact of a cost-recovery policy is to limit access. Only those individuals, agencies, and companies that can afford the data can access and use it. As described in "Special Issues on Access and Participatory Approaches" in the *Journal of the Urban and Regional Information Systems Association* (Onsrud and Craglia 2003), this raises a number of questions about democratic participation in public decisions using those data as well as concerns about asking members of the public to pay twice for the data.

The secondary impact was to limit or openly prevent the distribution and development of other geospatial data. In this example, a licensed user of the data used the parcel data as an important input for the development of its land-use/cover datasets. However, the boundaries of the land use/cover were determined to reveal too many details of the licensed parcel data and thus could only be distributed to other organizations possessing the parcel data license. A creative solution was found by which the agency produced a degraded raster version of the data that could be released, but it meant that the public was receiving different information than was actually used for public decisions by the agency.

There also exists a tertiary level of impact with significance for the GIS industry. Many of the surrounding municipalities have found it necessary to buy into the license for the previously mentioned data (either for the parcels or for the spin-off datasets). Those municipalities require bidders on a variety of GIS and engineering projects to acquire a data license so that they can use the municipal datasets. As a result, a local firm with an existing license (affordable when working on multiple licensed data projects across the region) has an upper hand over firms from out of town when competing for a local contract. Over time, this policy could impede competition and create an ad hoc good-old-boys network that reinforces the policy.

Concern about the ripple effect comes in the combined impact of the hidden ripples (shown in Figure 4). It probably wouldn't surprise most to hear that access restrictions in a heavy centralized model could impact GIS users throughout the entire community. But the previous example comes from a federated network where one actor significantly impacts many others within the network.

Costs of Enforcement

A repeated theme surfacing in interviews was the difficulty in constructing and enforcing a policy that could appropriately limit the redistribution of public geospatial data. Discussions with agencies engaged in either highly restrictive policies or price schemes very often lead to descriptions of complicated processes relying significantly on the work of a legal staff, diligent efforts to sniff out data poachers violating license agreements, or processes through which only known and trusted entities could access public data. In some cases, the fees charged didn't seem sufficient to cover the costs of establishing and enforcing the restrictions; in others, the enforcement was minimized to the status of being a major distraction [K3, K7].

The Benefits of Sharing

Another recurring theme was the value that many communities saw in sharing data. One interviewee was particularly ecstatic in stressing how important it was to simply serve the community. Others, however, saw significant political gains resulting from their open data-sharing policies. The ability to leverage the "free" data that they shared was seen as a politically expedient reason to continue sharing their data [C1]. At times this was done with a clear expectation of getting something in return, whether that was data from other organizations or simply a general recognition of the data development and dissemination effort that might lead to larger budget allocations.

The political expediency of data sharing depends on working to ensure that the agency of origination receives full and proper credit [C2] and careful and explicit communication about the costs of data development [C3]. One of the easiest ways to get credit is to make sure that data is accompanied by metadata that provides that information. At a time of budget cuts, a well-informed user community can be a much more important asset than a few thousand dollars in data-access fees. By sharing public geospatial data, interviewed data developers are able to impact other, sometimes critically important, technology applications [C5]. Many systems managers were very aware of the importance of their user community and worked to promote those applications in other departments (*e.g.,* a highly visible sheriff's department) more than their own data.

A different approach used by some communities has been to share their data as a way to promote their GIS development staff as a GIS Service Center. Charging a reasonable fee for services allows the agency to develop a revenue stream that is somewhat independent of budget cycles. It also allows the agency to develop exceptional examples of how their data can be applied [C4].

Formal/Informal Relationships

The most successful data-sharing networks clearly relied on a combination of formal and informal relationships to facilitate the largest distribution of spatial data. The problem with informal relationships is that sometimes they can be all about "who you know," which involves social concepts including trust (Harvey 2003) [K6]. A frequent complaint associated with formalized networks and sharing arrangements was the difficulty in negotiating and maintaining the arrangements and the rigidity that these arrangements often imposed on day-to-day functions. Much of the conflict between formal and informal is also related to the conflicts between administrative and political short-term and long-term conflicts and between personal and institutional goals.

Data Sharing and Metadata

Whether formal or informal, the development and exchange of metadata represented one of the most important practices in determining the success of large-scale data-sharing activities. While a variety of practices existed with respect to metadata, it was clear that networks with significant amounts of data sharing relied heavily on a well-developed system of documenting data. Some smaller data-sharing networks function without any notable form of metadata, relying instead on verbal communication and interpersonal relationships—strengthening the authority and power of persons controlling information about the data available for sharing. Larger organizations or networks rely much more heavily on the institutionalization of processes ensuring the development of metadata—whether formal, informal, or FGDC-compliant—because of the inability to provide support to each individual and organization that tries to use their data [K4]. The other interesting characteristic related to the relationship between data sharing and metadata is that organizations that took the effort to produce relatively detailed metadata almost always had geospatial data of some value. [K5]

GIS as a Social Phenomenon

Data-sharing networks in local government are made up of people and the relationships between those people. Any system relying on people and relationships is subject to the complexities of social coordination and the influences guiding or motivating involved individuals. One of the most difficult factors to address or change may be the fundamental beliefs or attitudes of individuals with authority over decisions regarding data sharing [K2]. This research encountered a wide variety of attitudes regarding data sharing, ranging from individuals who felt that free and open data sharing was a moral imperative to others who viewed it as a major burden of little personal importance to others who felt obliged to capitalize on their data in any way possible (particularly by charging significant access fees). While some of these perspectives can be explained by a particular context, they are also impacted by the personal attitudes of the individuals involved.

CONCLUSIONS

Best Practice Number 1: Education

The cases studies made clear the importance of education as the highest priority for data-sharing best practices that will make any major SDI initiative a success. Workshops or other educational meetings are one avenue, but written materials for local and regional governments would serve as invaluable resources. An "SDI Guide to GI Sharing and Coordination" obviously could complement workshops, but, more important, a guidebook would have its own legs and reach people and places that workshops barely scratch. In particular, the guide should address "data themes" that are most impacted and recommend that they focus on trying to "loosen them up" a little somehow (locally, specific themes such as parcels and zoning that would be hardest to develop top-down, for sharing policies often prevent these from being stitched together). Minigrants to generate contributions would be helpful; this is something that national organizations could coordinate. Attention to financial and copyright issues is critical. A stronger connection to geospatial activities and resources in national government is also needed.

In terms of guidance for SDI development in the United States, the *Framework Introduction and Guide* (FGDC 1997) can provide some preliminary ideas, although ten years later, much has changed and the need for new guidelines is evident.

Research Issues

Regarding research, two separate but related areas should be prioritized. First, a new and improved framework survey is urgently needed. We have too little information about what is occurring nationwide. Are people involved in building framework layers? Are they participating in the development of the U.S. NSDI? What strategies are people employing locally and regionally? Second, although our research turned up some valuable insights into the processes of data sharing and coordination, we were not able to spend enough time to actually find out how formal approaches interact with existing formal and informal approaches. In the United States, the NSDI isn't being built on a blank slate; most SDIs aren't. Every government agency has existing mandates that necessitated data production, sharing, or coordination in the past. We could hypothesize that the largest hindrance for the NSDI at the local level arises from missed chances to piggyback NSDI development on existing government policies and programs. The policy-level issues are many times abstract. The rubber hits the road in local government and we still don't understand how GI is practically interwoven in mandated and legislated activities.

Connected to this point is an equally urgent need to assess the financing models for GI in local governments. We have larger studies that elaborate the capitalization approach for building a national infrastructure, but what about the current financing models? Related questions about costs and copyright also are important. Finally, scholarship and debate surrounding the NSDI have quieted somewhat in recent years and should be reenergized. The efforts supporting a U.S. NSDI have benefited historically from regular prodding of Mapping Science Committee reports (1993, 1997, 2001), external examinations of the potential for an NSDI (*e.g.,* National Academy of Public Administration 1998), and reviews of the successes and failures of the program (Tosta 1999).

Organizations and Coalitions

The formal development of a coalition of organizations could help drive home the importance of data sharing for local governments and SDI development. URISA's Summit of Partnerships and Collaboration (http://www.urisa.org/FedSummit/Summit.htm) demonstrates that a variety of organizations are willing to rally around this topic area. The National Association of County Officials (NACO) and the League of Cities also have GIS-related activities that reach local governments. Obviously, other state and national groups should be involved. Small, but symbolically significant, awards for local governments could help boost awareness and make NSDI participation a more prestigious attribute.

Moving On

Best practices of data sharing and coordination involve many aspects of administrative and political activities. The best practices for any particular locality at any particular time are contingent on a number of factors and characteristics. Of all the practices we have identified in this research, a key practice seems to lie in the approach towards data and colleagues. If data sharing and coordination is just about the data, it will be very difficult at best, and may likely not work at all, nor for any length of time. Data sharing and coordination are best understood as part of other activities. Some of these activities require interaction; many others are assisted and promoted by data sharing. Establishing and supporting a social network among colleagues, citizens, and elected officials that supports their interests seems to be critical in all cases.

Like many institutional and social interactions, the critically important process of data sharing will continue to be mildly confounding but can be advanced by an awareness of the issues relating to data sharing by all involved parties. The issues described in this paper, like context and attitudes, contribute in different ways, placing an unexpected burden on many GI science professionals. Technology helps overcome some barriers, but it also creates new barriers. Similarly, social networks can be of great assistance, but they can also undermine individual efforts. Perhaps one the greatest long-term steps towards improving these conditions is a more thorough integration of ideas about data sharing, like the keys and contradictions, into educational curricula.

About the Authors

David Tulloch is an Associate Professor in the Department of Landscape Architecture at Rutgers, The State University of New Jersey. He also serves as the Associate Director of the Grant F. Walton Center for Remote Sensing and Spatial Analysis where he conducts PPGIS and Institutional GIS research and teaches in the Environmental Geomatics program.

Corresponding Address:
Grant F. Walton Center for Remote Sensing and Spatial Analysis
Rutgers, The State University of New Jersey
New Brunswick, NJ 08901 USA
dtulloch@crssa.rutgers.edu
732.932.1581p
732.932.1940f

Francis Harvey is an Associate Professor at the University of Minnesota, teaching in the MGIS and geography programs. His research focuses on organizational dimensions of GIS and SDI and includes long- term study of cadastral systems in relationship to land tenure.

Corresponding Address:
Francis Harvey
Department of Geography
University of Minnesota
Minneapolis, MN 55455
USA
fharvey@umn.edu

References

Azad, B., and L. L. Wiggins. 1995. Dynamics of inter-organizational geographic data sharing: a conceptual framework for research. In Onsrud H. J., and G. Rushton, Eds., Sharing geographic information. New Brunswick, NJ: Center for Urban Policy Research, Rutgers, The State University of New Jersey, 22-43.

Calkins, H. W., and R. Weatherbe. 1995. Taxonomy of spatial data sharing. In Onsrud, H. J., and G. Rushton, Eds., Sharing geographic information. New Brunswick, NJ: Center for Urban Policy Research, Rutgers, The State University of New Jersey, Chapter 4.

Craig, W. 1995. Why we can't share data: institutional inertia. In Onsrud, H. J., and G. Rushton, Eds., Sharing geographic information. New Brunswick, NJ: Center for Urban Policy Research, Rutgers, The State University of New Jersey, Chapter 6.

Epstein, E. 1991. In my opinion. Journal of the Urban and Regional Information Systems Association 3:2-4.

Federal Geographic Data Committee. 1997. Framework introduction and guide. Washington, D.C.: Federal Geographic Data Committee.

Harvey, F. 2003. Developing geographic information infrastructures for local government: The role of trust. Canadian Geographer 47(1): 28-37.

Harvey, F., and D. Tulloch. 2003. Building the NSDI at the base: establishing best sharing and coordination practices among local governments (report). Minneapolis, MN, and New Brunswick, NJ: University of Minnesota. June 2004, http://www.tc.umn.edu/~fharvey/research/BestPrac7-03.pdf

Harvey, F. and D. Tulloch. 2004. How do local governments share and coordinate geographic information? Issues in the United States. 10th EC-GI & GIS Workshop—ESDI: The State of the Art, Warsaw, Poland, June 23-25, 2004.

Harvey, F., and D. Tulloch. 2006. Local government data sharing: evaluating the foundations of spatial data sharing infrastructures. International Journal of Geographical Information Systems. 20(7): 743-68.

Kevany, M. 1995. A Proposed Structure for Observing Data Sharing. In Onsrud, H. J., and G. Rushton, Eds., Sharing geographic information. New Brunswick, NJ: Center for Urban Policy Research, Rutgers, The State University of New Jersey, Chapter 6.

Mackaay, E. 1982. Economics of information and law. Boston: Kluwer-Nijhoff Publishing.

Mapping Science Committee. 1993. Toward a coordinated spatial data infrastructure for the nation. Washington, D.C.: National Academy Press.

Mapping Science Committee. 1997. The future of spatial data and society: Summary of a workshop. Washington, D.C.: National Academy Press.

Mapping Science Committee. 2001. National spatial data infrastructure for the nation. Washington, D.C.: National Academy Press.

National Academy of Public Administration. 1998. Geographic information for the 21st century: building a strategy for the nation. A report by a panel of the National Academy of Public Administration for the Bureau of Land Management, Forest Service, U.S. Geological Survey, and National Ocean Service. Washington, D.C.: National Academy of Public Administration.

Nedovic-Budic, Z., J. Pinto, and L. Warnecke. 2004. GIS database development and exchange: interaction mechanisms and motivations. Journal of the Urban and Regional Information Systems Association 16(1): 15-29.

Onsrud, H. J., and G. Rushton. 1995. Sharing geographic information: an introduction. In Onsrud, H. J., and G. Rushton, Eds., Sharing geographic information. New Brunswick, NJ: Center for Urban Policy Research, Rutgers, The State University of New Jersey, Chapter 4.

Onsrud, H. J., and M. Craglia. 2003. Special issues on access and participatory approaches (APA) Number 1 and 2. Journal of the Urban and Regional Information Systems Association 15(APA 1 and 2).

Patton, M. Q. 2001. Evaluation, knowledge management, best practices, and high quality lessons learned. American Journal of Evaluation 22(3): 239-336.

Peters, T. , and R. Waterman, Jr. 1982. In search of excellence: lessons from America's best-run companies. New York: Warner Books, 1982.

Tosta, N. 1999. NSDI was supposed to be a verb: a personal perspective on progress in the evolution of the U.S. National Spatial Data Infrastructure. In Gittings, B., Ed., Integrating information infrastructures with GI technology: innovations in GIS 6. Philadelphia: Taylor and Francis, Chapter 2.

Tulloch, D. L., and E. Epstein. 2002. Benefits of community MPLIS: efficiency, effectiveness, and equity. Transactions in Geographic Information Systems 6(2): 195-212.

Watts, D. 2003. Six degrees. New York: Norton & Company.

When de Montalvo, U. 2003. In search of rigorous models for policy-oriented research: a behavioral approach to spatial data sharing. Special Issue on Access and Participatory Approaches (APA). Journal of the Urban and Regional Information Systems Association 15(1): 19-28.