Contents

Referred

5 NSDI Building Blocks: Regional GIS in the United States
Zerica Nedović-Budić, Gerrit-Jan Knaap, Nama Raj Budathoki, and Branko Cavrić

25 Hierarchical Clustering of Multiobjective Optimization Results to Inform Land-use Decision Making
C. M. Moulton, S. A. Roberts, and P. H. Calamai

39 The Potential of Integrating E-participation in Planning Support Systems
Arjen Koekoek, R. van Lammeren and G. Vonk

49 Using GIS to Contrast Perceived Versus Preferred Priorities for Brownfield Redevelopment in Worcester, Massachusetts
Claire W. Brill

Note: The Koekoek and Brill papers were prize winners in the 2008 URISA Student Paper Competition.
EDITORS AND REVIEW BOARD

URISA Journal Editor

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

Article Review Board

Peggy Agouris, Center for Earth Observing and Space Research, George Mason University, Virginia

David Arctur, Open Geospatial Consortium

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)

Itzhak Benenson, Department of Geography, Tel Aviv University (Israel)

Al Butler, Milepost Zero

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)

Paul Cote, Graduate School of Design, Harvard University

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

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

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

Michael Gould, Environmental Systems Research Institute

Klaus Greve, Department of Geography, University of Bonn (Germany)

Daniel A. Griffith, Geographic Information Sciences, University of Texas at Dallas

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

Richard Klosterman, Department of Geography and Planning, University of Akron

Jeremy Mennis, Department of Geography and Urban Studies, Temple University

Nancy von Meyer, Fairview Industries

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

Zorica Nedovic-Budic, School of Geography, Planning and Environmental Policy, University College, Dublin (Ireland)

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

Zhong-Ren Peng, Department of Urban and Regional Planning, University of Florida

Carl Reed, Open Geospatial Consortium

Claus Rinner, Department of Geography, Ryerson University (Canada)

Vonu Thakuriah, Department of Urban Planning and Policy, University of Illinois Chicago

Mary Tsui, Land Systems Group

David Tulloch, Department of Landscape Architecture, Rutgers University

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

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

Lyna Wiggins, Department of Planning, Rutgers University

F. Benjamin Zhan, Department of Geography, Texas State University-San Marcos
NSDI Building Blocks: Regional GIS in the United States

Zorica Nedović-Budić, Gerrit-Jan Knaap, Nama Raj Budhathoki, and Branko Cavrić

Abstract: This paper presents an assessment of the regional capacity in implementing geographic information systems (GIS) and databases. We draw on a Web survey to determine the availability and assembly of spatial data at the metropolitan level in the United States. Information was sought from 388 planning organizations and regional agencies located in 349 metropolitan areas. Based on 116 responses (30 percent response rate), we generate descriptive statistics and run a regression model addressing the following aspects of the regional GIS capacity: data (contents, update, and assembly); technology (compatibility of software and data formats, data access, and the use of standards); people (staff, leadership commitment and support, regional communication and cooperation); policy (data sharing, funding, rules and responsibilities, and mandates/programs); and context (urban and regional issues and affluence). Although the survey reveals only a snapshot of a dynamic and evolving phenomenon, the results indicate that the status of metropolitan GIS in the United States may not be matching what is technically feasible. While the capacity is getting better over time, the process is relatively slow and the challenges of creating the base for building the National Spatial Data Infrastructure (NSDI) are persistent. Future research and practice should place more emphasis on the relationship between the NSDI and its installed base. Nurturing of networks and compatibilities among organizational entities at various levels, local and regional in particular, should be given priority in devising policies and programs for a useful and sustainable spatial data infrastructure.

INTRODUCTION

Since their beginnings in the 1960s, digital geographic information systems (GIS)—used for collection, storage, management, and analysis of spatial data—have penetrated many societal segments and have established GIS as a scientific discipline (Longley et al. 2005). A survey by Public Technology, Inc., indicates that by the early 2000s, GIS have become integral resources in various local functions, including urban planning, public works, financial, public safety, and economic development (Public Technology, Inc. 2003). Both early (Budić 1993, Budić and Godschalk 1994, Crosswell 1991, French and Wiggins 1989, 1990) and recent GIS studies (Caron and Bédard 2002, Haithcoat et al. 2001, ICMA 2002, Kreizman 2002, NACo 1999, Norris and Demeter 1999, Warnecke et al. 1998) provide useful information about the extent of GIS diffusion and use in the United States. These studies are complemented by significant research on data sharing and interorganizational GIS—as the apparent consequences of proliferation of spatial databases in digital form and the related pressure to exchange and access information resources (Greenwald 2000; Harvey and Tulloch 2006; McDougall 2006; Montalvo 2003; Nedović-Budić and Pinto 1999a, 1999b, 2000, 2001; Nedović-Budić et al. 2004a; Omran 2007; Onsrud and Rushton 1995). However, there is not much research on data assembly and availability at the regional metropolitan level.

Why is the regional metropolitan level important? First, more than three-fourths of the U.S. population (U.S. Census Web site) lives in metropolitan regions where most of urban growth and its implications occur and where large numbers of people are vulnerable in disaster situations. Second, many societal problems are tackled in a more holistic and coordinated manner at this level (Alliance for Regional Stewardship Web site, Feiocock 2007, Wallis–MuniMall Web site, including the response to emergencies (Alliance for Regional Stewardship 2002). This is particularly true in the United States where the local level is the ultimate locus of decision making and action, but is extremely fragmented, often artificially bounded, and plagued by overlapping jurisdictions of more than 85,000 different local government entities (U.S. Census Bureau Web site). Third, regional GIS represent a base of the National Spatial Data Infrastructure (NSDI) (Georgiadou et al. 2006, Harvey and Tulloch 2006, Nedović-Budić and Budhathoki 2006, Rajabifard et al. 2006).

The U.S. NSDI is defined as “a physical, organizational, and virtual network designed to enable the development and sharing of this nation’s digital geographic information resources” (Federal Geographic Data Committee (FGDC) Web site). It is realized through the development of spatial data and metadata standards, the establishment of spatial data clearingshotes (Geospatial One Stop), and the identification of national data sets (so-called “framework data,” Tulloch and Fulld 2001), and more recently The National Map. Over the past decade, the NSDI concept has moved from a data-centered to a process-centered approach (Masser 2005), but the access to spatial data and information about available data has remained its primary tenet. In particular, the access to local and regional (i.e., subnational) data sets is recognized as the key element of the second generation of SDIs (Rajabifard et al. 2006). While the NSDI does not assume database assembly across the national territory, it does require data availability and compatibility to enable rapid and convenient data integration based on the user’s needs.

The research presented in this paper is concerned with the provision and integration of spatial data at the metropolitan level. The objectives are to (a) assess GIS data availability at the metropolitan level and (b) identify the factors associated with assembling and sharing GIS data across multiple jurisdictions within a metropoli-
The empirical evidence is acquired with a Web-based national survey of regional data collection and assembly efforts in 349 U.S. metropolitan areas. While the data does not reflect the most up to date status of regional GIS, it approximates a complex process somewhat stalled by institutional inertia, and is used to illustrate the challenge of building the NSDI base, particularly in a country as large as the U.S.

After the discussion of the research premises, frameworks, and methodology, we present the results based on descriptive statistics and a regression model. We conclude with a summary of the findings, implications for building a sustainable and viable NSDI, and suggestions for further research.

**LINKING DATA SHARING, INTERORGANIZATIONAL GIS, AND SDI**

**Interorganizational GIS and Information Infrastructure**

Fundamentally, regional GIS depend on developing seamless databases and techniques, securing incentives and resources for data integration, and establishing multijurisdictional intergovernmental cooperation. When geospatial technologies and information resources are distributed across organizational boundaries to include multiple local governments and nonprofit groups, or to involve private-sector partners, they form interorganizational GIS (O’Loan 1997). These systems draw on existing interdependencies, but also are challenged by their complexities (Nedović-Budić and Pinto 1999b). To frame the complexities involved in building distributed systems, Fletcher (1999) proposes four levels of interoperability: global, regional, enterprise, and product; three types of interoperability: institutional, procedural, and technical; and three dimensions of interoperability: horizontal, vertical, and temporal.

The most important factors for achieving interoperability and multiparticipant geospatial technologies and systems are sharing and easy access to geospatial information. Sharing geospatial information is believed to promote more effective use of organizational resources and cooperation among involved organizational entities (Brown et al. 1998, Nedović-Budić and Pinto 1999a, 1999b, 2001). Obstacles to data sharing are numerous, including both technical and nontechnical issues. On the technical side, for example, it is very hard to resolve the varying needs for scales and accuracy of data that users located in the same region may have. On the nontechnical side, there may be inadequate communication about available information resources or a lack of willingness to share those resources. These are the same factors and issues the information infrastructures (IIs) and spatial data infrastructures (SDIs) in particular are established to facilitate and resolve by introducing a mechanism for a diverse set of data producers and users to interact in an open networked environment. “Data sharing among the participants on an unprecedented scale will be needed for SDIs to become fully operational and effective in practice” (Rajabifard et al. 2006, p. 738).

The connection between interorganizational systems and IIs also is acknowledged in research literature. For example, while proposing the characteristics of IIs, Star and Ruhleder (1996) argue that they cannot be independently built and maintained, but, rather, they emerge through practice and become connected to other activities and structures. Similarly, Borgman (2000) views IIs as much more than the physical substrate and thus considers broader social relations in constituting IIs. Hanseth and Monteiro (1998) suggest that some of the II characteristics may be present in certain information systems (IS), especially in interorganizational systems (IOS) or distributed information system (DIS) and, therefore, some commonalities and overlapping characteristics exist between IS and II (see Table 1, Budhathoki and Nedović-Budić 2007). The authors consider that IIs are initiated when: new and independent actors become involved in the development of an IOS or DIS, so that the development is not controlled by one actor anymore; one of the design objectives for IOS or DIS is to grow and become an II (or part of an II) in the future.

**Local and Regional GIS as the NSDI Building Blocks**

We propose that the interorganizational GIS at the metropolitan regional level constitute an installed base and building blocks of the U.S. NSDI. NSDI is defined as “the technology, policies, criteria, standards and people necessary to promote geospatial data sharing throughout all levels of government, the private and non-profit sectors, and academia” (FGDC Web site). Rajabifard et al. (2006) consider spatial data infrastructure (SDI) “an enabling platform for data sharing” (p. 727). The authors differentiate between the first generation SDIs that are mainly led by national mapping agencies and focused on provision of national data sets, and the second generation SDIs that are process-based interactions between multiplicity of players in the joint effort toward managing and exchanging information assets in a networked environment. The switch from the first to the second generation happened around the year 2000 when a centralized (or top-down) product-oriented model was replaced by a bottom-up distributed model. Accordingly, the product-based model involves definition of data, collection of data, integration of data, database creation, and implementation; the process-based model involves knowledge infrastructure, capacity building, communication, and coordination. In relating the SDI hierarchy and the models, Rajabifard et al. (2003) associate SDIs from local to state level to the product model and operational tier; national SDIs with the management tier and both product and process models; and multinational1 and global SDIs with strategic tier and process model only. Obviously, availability of and accessibility to spatial data remain the core of a functional SDI, although the services are increasingly being added to SDI clearinghouses and portals (Crompvoets and Bregt 2007).

---

1 We substitute the term multinational for the term regional used by the authors to avoid confusion with the regional metropolitan level that this research is concerned with.
Table 1. Characteristics of information infrastructures

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embeddedness</td>
<td>“Infrastructure is “sunk” into (inside of) other structures, social arrangements and technologies.”</td>
</tr>
<tr>
<td>Transparency</td>
<td>“Infrastructure is transparent in use, in the sense that it does not have to be reinvented each time or assembled for each task, but invisibly support those tasks.”</td>
</tr>
<tr>
<td>Reach or scope</td>
<td>“This may be either spatial or temporal—infrastructure has reach beyond a single event or one-site practice.”</td>
</tr>
<tr>
<td>Learned as part of membership</td>
<td>“The taken-for-grantedness of artifacts and organizational arrangements is a sine qua non of membership in a community of practice. Strangers and outsiders encounter infrastructure as a target object to be learned about. As they become members, new participants acquire a naturalized familiarity with its objects.”</td>
</tr>
<tr>
<td>Links with conventions of practice</td>
<td>“Infrastructure both shapes and is shaped by the conventions of a community of practice.”</td>
</tr>
<tr>
<td>Embodiment of standards</td>
<td>“Modified by scope and often by conflicting conventions, infrastructure takes on transparency by plugging into other infrastructures and tools in a standardized fashion.”</td>
</tr>
<tr>
<td>Installed base</td>
<td>“Infrastructure does not grow de novo; it wrestles with the ‘inertia of the installed base’ and inherits strengths and limitations from that base.”</td>
</tr>
<tr>
<td>Becomes visible upon breakdown</td>
<td>“The normally invisible quality of working infrastructure becomes visible when it breaks.”</td>
</tr>
<tr>
<td>Hanseth and Monteiro (1998), pp. 41-49</td>
<td></td>
</tr>
<tr>
<td>Enabling</td>
<td>“Infrastructures have a supporting or enabling function.”</td>
</tr>
<tr>
<td>Shared</td>
<td>“An infrastructure is shared by a large community (collection of users and user groups).”</td>
</tr>
<tr>
<td>Open</td>
<td>“Infrastructures are open and support heterogeneous environments.”</td>
</tr>
<tr>
<td>Sociotechnical network</td>
<td>“IIs are more than ‘pure’ technology; rather, they are socio-technical networks.”</td>
</tr>
<tr>
<td>Ecology of networks</td>
<td>“Infrastructures are connected and interrelated, constituting ecologies of networks.”</td>
</tr>
<tr>
<td>Installed base</td>
<td>“Infrastructures develop through extending and improving the installed base.”</td>
</tr>
</tbody>
</table>

Local and regional (metropolitan) levels are the most relevant instances of spatial data production and data use and they could represent the building blocks of the NSDI (Nedović-Budić and Budhathoki 2006, Rajabifard et al. 2006). However, the connectivity between these levels is not easy to operationalize. For example, Harvey and Tulloch (2006) consider the available databases and the uptake of SDI concepts at these levels as essential for the success of the U.S. NSDI, but their 2002 case study shows that majority of U.S. local governments are either unaware of or do not take the SDI concept relevant to them. This is reinforced in a study by Nedović-Budić et al. (2004b) who report that the state SDI in Illinois does not meet the needs of local government planners. Similarly, in the emergency context, despite the important supporting role GIS has played in responding to 9-11 (New York 2001) and Katrina (New Orleans 2006) crises, the challenge of quick data integration that could be provided only by a viable SDI is recognized (Adam et al. 2006, Butler 2005). Agreeing with these findings, Dresler and Woods (2000), in a summary of six community demonstration projects supported by the FGDC, point out the advantages and shortcomings of the federal SDI-related activities. Beside numerous positive developments, they report that “[i]nformation required to address very localized issues such as growth, flooding, and crime analysis often require higher resolution data than is presently collected by the Federal community” (p. 6). Finally, despite major efforts and achievements in building NSDIs, based on their global assessment, Crompvoets et al. (2004) observe a declining trend of clearinghouse use and suggest that user-unfriendly interface and discipline-specific nature of metadata and clearinghouses are among the primary reasons for the decline. Clearly, the previous statements reafirm the calls for the development of bottom-up NSDIs rooted in subnational governments and based on the needs of spatial data users (Budhathoki et al. 2008, Rajabifard et al. 2006).

U.S. National Spatial Data Infrastructure

The United States is among the first countries to embrace the idea of building the National Spatial Data Infrastructure (NSDI) since the early 1990s. The main impetus is given by President Clinton’s Executive Order 12906 of April 1994 (FGDC Web site) and the Office of Management and Budget’s (OMB) Circular A-16 and E-government Act of 2002 (FGDC Web site, U.S. OMB Web site). Development of national spatial data infrastructures has been undertaken in many countries worldwide. In their longitudinal survey, Crompvoets and Bregt (2007) report that by 2005 83 countries have adopted national-level SDI programs. SDIs also are developed at other levels, such as regional, state, and local (Rajabifard et al. 2002, Masser 2005), and billions of dollars are spent each year on SDI-related activities worldwide (Onsrud et al. 2004).

Substantial progress has been made since the inception of the NSDI in the United States. Following the efforts in conceptualizing the NSDI, there have been numerous activities in development of data and metadata standards, awareness-raising activities at all levels, establishment of clearinghouses, definition of framework data, and creation of partnerships to facilitate spatial data availability and access (FGDC Web site). The importance of building a sustainable and useful NSDI became particularly apparent after the events of 9-11 (CAD Digest Web site, GIS Monitor Web site) and Hurricane Katrina (UCGIS Web site). In addition to the Federal Geographic Data Committee, the U.S. NSDI is building on two other initiatives—the National Map (USGS Web site) and Geospatial One Stop (Geodata.gov Web site), which are all brought together by
the National Geospatial Program Office. Also the Department of Homeland Security has restructured to better handle geospatial information of national interest (National Geospatial-Intelligence Agency Web site, ESRI Web site).

Despite these efforts and strategic documents issued in 1994, 1997, and 2004 (FGDC Web site), the U.S. NSDI still is not supported by a comprehensive and operational implementation plan. The latest NSDI Future Directions Initiative (2004) is a vague guiding document that lacks the programming component. Past efforts have been focused primarily on the federal level where the standardization activity is mandatory, but where full coordination still is missing. The main NSDI building tools are data partnerships. Thus, the local and state levels have been tackled through partnerships with national associations (e.g., the National State Geographic Information Council (NSGIC), the National Association of Counties (NACo), and the International City/County Management Association (ICMA)), as well as through direct contacts with government organizations at all levels. The FGDC Cooperative Agreement Program (CAP) that has been operating since the mid-1990s generates many seed projects and test beds of NSDI implementation. However, these CAP-supported projects could not amount to a nationally significant outcome (Mapping Science Committee 2001) or reach the organizations most in need for funding (MacPherson et al. 2003). The Mapping Science Committee (2001) finds that “funding incentives established by the FGDC through the NSDI partnership programs do not appear to have significantly reduced data redundancy, decreased cost, improved access, and increased accuracy.

The states have been approached through the NSGIC with the 50 States Initiative (FGDC 50 States Initiative Web site), primarily for drafting the strategic and business plans. These plans are to “facilitate the coordination of programs, policies, technologies, and resources that enable the coordination, collection, documentation, discovery, distribution, exchange and maintenance of geospatial information in support of the NSDI” (50 States Initiative Web site). By October of 2007, strategic plans are completed in nine states, pending in four, in progress in ten, in final draft in one, starting in eight, N/A in one, and unknown in four; business plans are completed in seven states, pending in four, in progress in ten, starting in eight, N/A in three, and unknown in five states. CAP funding is used to support the development of strategic and business plans, but there are no other NSDI implementation resources committed. Otherwise, regardless of NSDI-related initiatives, GIS coordination at the state level is present in many of the 50 states (Warncke et al. 2003).

Finally, the regions, as probably the most viable link between the local, state, and federal levels in the U.S. NSDI, are somewhat neglected. Suitability of the regional level as data assembly and distribution point, in particular, has been acknowledged early in the conceptualization of the U.S. NSDI through the idea of “area integrators” (FGDC 1995). Unfortunately, this idea never was implemented and the opportunity to build an NSDI with a strong regional and local base was missed. There is currently a revival of this idea through the National Geospatial Advisory Committee (NGAC) established in January of 2008. The NGAC reports to the FGDC chair and “provide[s] a forum to convey views representative of non-federal stakeholders in the geospatial community” (NGAC Charter, NGAC Web site). Local and regional governments and organizations are among the most important nonfederal stakeholders.

Comprehensive studies of the U.S. NSDI are scarce. Following the 1998–1999 Framework Survey, an inventory of organizations that produce or use framework data, availability of metadata, data sharing practices, and key contacts (Harvey 2001, Tulloch and Robinson 2000, Tulloch and Fuld 2001), there has been no systematic nationwide attempt to find out about the status of the U.S. NSDI. The Framework Survey suggests that the use of framework data in an SDI environment is challenging technically and institutionally: technically because data are in various formats and of different accuracies; institutionally because data producers are not fully prepared to share data. Other research efforts, including the one presented here, address the phenomena relevant to building the NSDI—policy and organizations; interoperability and sharing; and discovery, access, and use of spatial data (Budhathoki and Nedović-Budić 2007). The following sections discuss the research framework and the empirical evidence on the regional GIS capacity as the installed base and building block of the U.S. NSDI.

RESEARCH FRAMEWORK AND METHODOLOGY

Conceptual Framework
This assessment of regional GIS capacity is conceptually related to the literature and research frameworks on GIS, interorganizational GIS, and SDI (see Table 2). The presentation of findings is organized around five broad concepts that are commonly featured across the three frameworks: data, technology, people, policy, and context. This study builds directly on the research conducted by Nedović-Budić and Pinto (Nedović-Budić and Pinto 1999a, 199b, 2000, 2001; Nedović-Budić et al. 2004a) and includes variables that they discover as important in the process of data sharing and building interorganizational and multijurisdictional GIS. Indirectly, the research also captures Fletcher’s (1999) regional and product levels of interoperability, procedural and technical type of interoperability, and horizontal and temporal dimensions of interoperability.

The primary focus of this study is the spatial (GIS) data held by governments in metropolitan areas. Inquiry into data available in individual jurisdictions (contents), data currency (update), and data assembly across jurisdictional boundaries are of main concern. In addition to finding out if specific data layers are available in some or all units (counties, parishes, or boroughs) that make up each metropolitan area, we are interested in identifying layers with regional coverage. We consider a comprehensive set of data themes, primarily because this work is driven by urban and regional planning concerns, which require such diversity and integration of spatial data (Dandekar 1988, Klosterman 2000). This diversity generally is found at the local government level, as reported in a U.S.
The authors find that more than 40 percent of the local governments sampled have the following components in their geospatial database: roads, hydrology, political/administrative boundaries, cadastral/land records, land use/zoning, elevation, digital imagery, and geodetic control. Those layers indicate the common data needs at the local level. With the addition of fire, police, and medical facility information, these local databases can easily meet the requirements of emergency applications as well.

Regional GIS capacity is affected by the *technological* factors such as compatibility of software used by various jurisdictions in the region, related data formats and map projections, data access methods used by organizations in the region, as well as the application of common data standards. It also is influenced by the nontechnological factors, including human (staffing, leadership commitment and support, regional communication and cooperation), policy (data sharing, formalization of data-related activities, funding of regional GIS, definition of roles and responsibilities, mandates/programs), and contextual (size of the region measured in the number of units it consists of, the region's affluence measured in per-capita income, and the presence of urban and regional issues such as sprawl and natural resources that could stimulate regional actions). Relevance of these factors is confirmed in GIS (Croswell 1991, Budić 1993) and interorganizational GIS sources (Nedović-Budić and Pinto 1999a, 1999b, 2000, 2001; Onsrud and Rushton 1995), but also has transferred into the SDI field (Askew et al. 2005, Craig 2000, Georgiadou et al. 2005, Tait 2005).

**Table 2. Conceptual frameworks used to inform the study**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FACTORS - TECHNOLOGY:</strong></td>
<td><strong>FACTORS - DATA:</strong></td>
<td><strong>FACTORS - PEOPLE:</strong></td>
<td><strong>FACTORS - POLICY:</strong></td>
</tr>
<tr>
<td>- compatibility of software</td>
<td>- contents</td>
<td>- staff</td>
<td>- data sharing/exchange activities</td>
</tr>
<tr>
<td>- compatibility of data formats and map projections</td>
<td>- update</td>
<td>- leadership commitment, support, and cooperation</td>
<td>- formalization</td>
</tr>
<tr>
<td>- methods of data access</td>
<td>- assembly across the region</td>
<td>- regional communication and cooperation</td>
<td>- rules and responsibilities</td>
</tr>
<tr>
<td><strong>FACTORS - Context:</strong></td>
<td><strong>FACTORS - PEOPLE:</strong></td>
<td><strong>FACTORS - POLICY:</strong></td>
<td><strong>FACTORS - CONTEXT:</strong></td>
</tr>
<tr>
<td></td>
<td>- contents</td>
<td>- data sharing/exchange activities</td>
<td>- regional per capita income</td>
</tr>
<tr>
<td></td>
<td>- update</td>
<td>- formalization</td>
<td>- number of counties in a region (size)</td>
</tr>
<tr>
<td></td>
<td>- assembly across the region</td>
<td>- rules and responsibilities</td>
<td>- planning issues (urban sprawl, natural resources)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sampling and Data Collection**

Metropolitan planning organizations (MPOs) are basically the only public sector regional organizations within the U.S. institutional structure that deal with and integrate spatial data. Local governments have statutory authority to perform administrative functions and enact and implement programs and policies. They are the loci of urban decisions and activities. Regions in the United States, however, do not have such powers, despite the fact that they represent a more appropriate unit of policy and action in many domains, urban and regional planning and emergency response, in particular.2 MPOs are established in each metropolitan area (MA) as a condition for receiving federal highway or transit funds in urbanized areas (Association of MPO Web site). The MPOs have responsibility for planning, programming, and coordinating federal highway and transit investments.

2 Portland Metro is the only constituted regional government in the United States (http://www.metro-region.org/).
Most MPOs have a larger scope of activities, including land-use analysis and planning as one of their standard tasks closely tied to transportation planning and modeling. Many MPOs tackle a variety of other urban and regional issues and rely on sophisticated methods and GIS databases (Greenwald 2000). Regional councils, commissions, and associations of governments are region-specific and also address a wide range of planning concerns. They are not required by law, but are frequently established in urban areas to assist community leaders and citizens in developing strategies for attending to transportation, economic development, air and water quality, social equity, growth, housing, and other urban and regional challenges. Increasingly, those challenges include management of emergency situations. In many cases, regional councils or commissions also are the official MPOs.

With their institutional nature and activities, MPOs and other regional organizations provide a conceptual and practical tie between the enabling function of information infrastructures and their link to communities of practice (Hanseth and Monteiro 1998, Star and Ruhleder 1996). The primary field of practice incorporated in this study is urban and regional planning, as probed by Nedović-Budić et al. (2004b) in their evaluation of utility of state-level SDI for local applications. Other authors also mention urban and regional planning as one of the main justifications of SDIs (Craglia and Johnston 2004, Masser 2005, NRC 1993).

In 2001, the U.S. Census Bureau identified 349 metropolitan areas (MAs), including: 261 metropolitan statistical areas (MSAs); 76 primary metropolitan statistical areas (PMSAs — aggregated in 19 consolidated metropolitan statistical areas or CMSAs); and 12 New England consolidated metropolitan areas (NECMA). Six MAs and one CMSA are located in Puerto Rico (U.S. Census Bureau Web site).3 Corresponding to the structure of their respective metropolitan areas MPOs’ jurisdictions range from one to over 10 constituent units (counties, parishes or boroughs).

A Web survey was used to collect data from the contacted organizations (shown in Figure 1). The Web form was developed internally by the UIUC Department of Urban and Regional Planning using Macromedia DreamWeaver MX and custom coding (http://www.urban.illinois.edu/faculty/budic/W-metroGIS.htm). Once submitted, the responses were written to a Microsoft Access 2000 database. Data was collected from July of 2002 to March of 2003.3 The sample for this study drew from the lists of metropolitan planning organizations and other regional agencies dealing with planning issues and located in one of the U.S. metropolitan areas. The surveys were filled out by staff members in managerial positions who were aware of (and in some cases involved in) regional GIS activities.

Responses were received from 116 agencies, or 30 percent of all metropolitan organizations contacted. The initial wave of responses was received during the summer and fall of 2002 and included 64 organizations. Most of them (61 or 95.3 percent) were the official metropolitan planning organizations located across 34 states. These 64 responses constituted a random cross section of metropolitan organizations that were used in further data analysis. This initial set of responses is referred to as “SAMPLE 64.” Additional responses were obtained in the fall of 2002 and the spring of 2003 after a targeted solicitation of responses from large urban areas by telephone and e-mail. Because the second wave of solicitation did not include all nonrespondents and the element of randomness was lost, the responses from the large urban areas were analyzed separately from the initial sample. The second set was a targeted sample consisting of 49 responses received from organizations located in the top 50 most populated urban areas in the United States—14 were eligible from the set of initial 64 respondents (i.e., 14 responses were from top 50 urban areas) and 35 came from the second set of respondents. The majority of these 49 responses (42 or 85.7 percent) were the official MPOs located in 29 states. In the remainder of this report, this set is referenced as “TOP 50/49.”

The 64 initial responses (SAMPLE 64) and the 49 responses from large urban areas (TOP 50/49) were nationally well dispersed (shown in Tables 3a and 3b and Figure 2). With regard to metropolitan area size, the 64 initial responses were a representative cross section of metropolitan areas, i.e., the percentage of respondents in each size category corresponded closely to the distribution of all metropolitan areas in those categories. Almost 90 percent of the 49 responses from the largest 50 urban areas were from metropolitan areas of more than one million inhabitants. The SAMPLE 64 offered insights into the national trends; the TOP 50/49 responses

Figure 1. The Web survey form

3 The source of MPO addresses and contacts was the 2001 Profile of Metropolitan Planning Organizations (Association of Metropolitan Planning Organizations 2001 Web site). The National Association of Regional Councils (NARC Web site) provided addresses of other regional organizations (e.g., councils, commissions, and associations of government). To secure information for each metropolitan area, both the metropolitan planning organizations and the regional councils/commissions were contacted with a request to fill in and submit the Web form. A total of 388 organizations were contacted—374 MPOs and 14 other regional organizations.

4 Even with the time passed between data collection and this reporting, the situation it portrays has not changed substantially. The rapid technological advances are followed by much slower institutional change and their full incorporation and use usually lag behind the potentials (Budić and Godschalk 1994, Campbell and Masser 1995, Masser and Onsrud 1993).
pointed to specific circumstances and issues in the most populous metropolitan areas.

One concern with the sampling was the unavoidable bias introduced by the respondents’ self-selection. The respondents in the random SAMPLE 64 could be described as more interested in and concerned with regional spatial data sharing and integration than the nonrespondents and, consequently, the assessment would be on the more optimistic side. In fact, many respondents were involved in regional GIS activities, primarily as coordinators or members—about one-third in each group from the random SAMPLE 64 and about one-half and one-fifth, respectively, from the TOP 50/49 largest urban areas. Very few respondent organizations acted as financial sponsors, subscribers, or observers of the regional GIS activities. On one hand, the active participation status was a factor that could potentially influence the responses. On the other hand, this same factor could ensure more informed and accurate answers.

**FINDINGS**

**Data**

Availability of specific content data was determined for each metropolitan area according to the respondents’ awareness about data themes developed for their organizations’ basic constituent units (counties, parishes, or boroughs) and about assembling data themes across all constituent units. The survey provided for three possibilities:

- **NONE:** meaning that the specific data theme did not exist in any of the counties, parishes, or boroughs in the organization's region;
- **ALL:** meaning that the specific data theme existed for all counties, parishes, or boroughs in the organization's region; and
- **SOME:** meaning that the specific data theme was developed for some (but not all) of the counties, parishes, or boroughs in the organization's region.

Data categories included land, regulation, boundaries, natural features, transportation, utilities, services, utilities, and

<p>| Table 3a. Distribution of responses in SAMPLE 64 by census region and area population |
|----------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Region</th>
<th>SAMPLE 64 Number</th>
<th>SAMPLE 64 Percent</th>
<th>MAs Number</th>
<th>MAs Percent</th>
<th>Population Under 500,000 Number</th>
<th>Population Under 500,000 Percent</th>
<th>Population 500,000-1 million Number</th>
<th>Population 500,000-1 million Percent</th>
<th>Population 1-5 million Number</th>
<th>Population 1-5 million Percent</th>
<th>Population Over 5 million Number</th>
<th>Population Over 5 million Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>10</td>
<td>15.6%</td>
<td>72</td>
<td>20.6%</td>
<td>37</td>
<td>57.8%</td>
<td>235</td>
<td>68.5%</td>
<td>11</td>
<td>17.2%</td>
<td>44</td>
<td>12.8%</td>
</tr>
<tr>
<td>Midwest</td>
<td>19</td>
<td>29.7%</td>
<td>82</td>
<td>23.4%</td>
<td>11</td>
<td>17.2%</td>
<td>44</td>
<td>12.8%</td>
<td>13</td>
<td>20.3%</td>
<td>59</td>
<td>17.2%</td>
</tr>
<tr>
<td>South</td>
<td>21</td>
<td>32.8%</td>
<td>131</td>
<td>37.4%</td>
<td>3</td>
<td>4.7%</td>
<td>5</td>
<td>1.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>14</td>
<td>21.9%</td>
<td>65</td>
<td>18.6%</td>
<td>Total</td>
<td>100.0%</td>
<td>Total</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>100.0%</td>
<td>350*</td>
<td>100.0%</td>
<td>Total</td>
<td>100.0%</td>
<td>Total</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*This count excludes 6 MAs located in Puerto Rico and double counts for 7 MAs located in two regions.

**Table 3b. Distribution of responses in TOP 50/49 by census region and area population**

<table>
<thead>
<tr>
<th>Region</th>
<th>SAMPLE 50/49 Number</th>
<th>SAMPLE 50/49 Percent</th>
<th>TOP 50 Urban Areas Number</th>
<th>TOP 50 Urban Areas Percent</th>
<th>Population Under 500,000 Number</th>
<th>Population Under 500,000 Percent</th>
<th>Population 500,000-1 million Number</th>
<th>Population 500,000-1 million Percent</th>
<th>Population 1-5 million Number</th>
<th>Population 1-5 million Percent</th>
<th>Population Over 5 million Number</th>
<th>Population Over 5 million Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>5</td>
<td>10.2%</td>
<td>6</td>
<td>12.0%</td>
<td>0</td>
<td>0.0%</td>
<td>1</td>
<td>2.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>11</td>
<td>22.4%</td>
<td>11</td>
<td>22.0%</td>
<td>5</td>
<td>10.2%</td>
<td>12</td>
<td>24.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>21</td>
<td>42.9%</td>
<td>20</td>
<td>40.0%</td>
<td>38</td>
<td>77.6%</td>
<td>32</td>
<td>64.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>12</td>
<td>24.5%</td>
<td>13</td>
<td>26.0%</td>
<td>6</td>
<td>12.2%</td>
<td>5</td>
<td>10.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>100.0%</td>
<td>50</td>
<td>100.0%</td>
<td>Total</td>
<td>100.0%</td>
<td>Total</td>
<td>100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*This count excludes 6 MAs located in Puerto Rico.

**Figure 2. Spatial distribution of SAMPLE 64 responses**
other (satellite imagery, aerial photography, and real estate/market data). Data was considered assembled if the theme data for the entire region was integrated into a single data set. Therefore, it was most interesting to find out to what extent the data was available and assembled across ALL constituent units. In SAMPLE 64, the most developed data themes were municipal boundaries, street/road network, traffic analysis zones (TAZs), and hydrology (rivers/streams), reported as available for ALL constituent units by 86 percent, 83 percent, 78 percent, and 78 percent of respondents, respectively. However, the same data sets were assembled by less than one-half of the regions that had data themes available across ALL units, and only about one-third of all respondents (see Appendix A).

Other data themes were less developed. For example, land use and parcel boundaries were available across ALL constituent units in 55 percent of the responses, but much less in their assembled form—44 percent and 30 percent, respectively. Inventory of developed, vacant land and infill sites, growth boundaries and service, zoning, and comprehensive plans were reported by about one-third of the organizations as available and about one-fifth as assembled. Finally, the percentages for data themes on urban infrastructure (e.g., sewer and water utilities) were considerably lower.

The respondents also reported the frequency of updating and assembling specific data themes. The options were monthly, semiannually, annually, every five years, every ten years, every ten+ years, and irregularly. The respondents were to provide an estimate that would represent an average across various constituent units holding a particular data theme. This was a difficult task that resulted in many missing responses. However, with a few exceptions, the responses received showed that “irregularly” was an overwhelming response across the themes for both data availability and data set assembly. Land use, land cover, and aerial photography were in many cases updated and assembled at least once in five years; tax assessment records, parcel boundaries, and TAZs were addressed at least annually, and the former bimonthly as well; finally, some of the service boundaries (e.g., schools) also were relatively regularly updated or assembled. Figures in Appendix B illustrate the geographic distribution of responses on availability and update of land use, street/road data, aerial photography, and satellite imagery for the SAMPLE 64 and TOP 50/49.

On average, data themes were more available across the constituent units and more likely to be assembled in the largest 50 urban areas than in the case of the random responses that came from metropolitan organizations located in urban areas of various sizes.

**Technology**

Compatibility of GIS software, formats, and map projections and access to spatial data used to be among the most challenging technical issues to resolve in multijurisdictional and multiorganizational settings. Over time, however, with the stabilization of the commercial software market, the development of conversion tools, and the availability of open codes, the software-related problems have decreased. The respondents to the Web survey indicated that, in general, most of the local governments in their regions used at least somewhat compatible software—56 (87.5 percent) responses from the random SAMPLE 64 and 41 (83.6 percent) responses from the TOP 50/49.

A range of methods of access to data developed and maintained by local governments in the region included Web mapping service (WMS), Internet-based access with automated file conversion system, centralized GIS storage/data warehouse, uploading/downloading via a computer network, and delivery via digital media (e.g., CDs, tapes, ZIP disks). Data access via portable digital media was the main method for a majority of the respondent organizations (see Figure 3). Other access methods also were present, with network-based access available to about half of the respondents and the Internet-based method only starting to emerge. Respondents from the TOP 50/49 urban areas reported a somewhat higher reliance on Internet or network-based approaches than the respondents from the random SAMPLE 64. However, the difference was not substantial.

The respondent organizations did not experience the access to data developed and/or maintained by other local organizations to be too difficult. More than two-thirds of the respondents (47 or 73.4 percent from SAMPLE 64 and 35 or 71.5 percent from the TOP 50/49) stated that the access was easy or somewhat easy, or were neutral about it. The respondent organizations from the TOP 50/49 largest urban areas found the access difficult or somewhat difficult slightly more often than the respondents across various sizes of metropolitan areas (15.7 percent versus 22.4 percent, respectively).

Using perceived difficulty in converting between map projections and formats as an indicator of facility of data integration, we found that data conversion did not represent an obstacle for the majority of respondent organizations. About two-thirds of them reported a low or very low level of difficulty with different GIS file formats, except for a somewhat higher percentage from the sample of 64 initial respondents reporting some difficulty with varying map projections used by local governments and other organizations in their region. No respondent qualified the conversion difficulty to be “very high.”

Finally, with respect to the use of standards, the responses pointed to limited efforts: 40 respondents (62.5 percent) from the random SAMPLE 64 group and 32 respondents (65.3 percent) from the TOP 50/49 group had no mutually accepted standards across the region.

**Policy**

Existing activities and mechanisms for regional data development and sharing were considered fundamental for securing access to metropolitan data sets and a step toward building the NSDI. Based on previous studies of interorganizational GIS, the following joint activities were considered: shared applications, shared geographic data clearinghouse, fee-based data exchange, coordinated database maintenance, shared database, coordinated data development and/or acquisition, and data sharing (free of charge). Free-of-charge data sharing as the most basic type of interaction was reported by the majority of respondents (shown in Figure 4). While this was encouraging and valuable, particularly for its open
nature, more complex interactions, including the establishment of clearinghouses, were pursued on average by less than one-fifth of the respondents. Consistent with responses on data availability and assembly, the existence of shared databases was sporadic, particularly among respondents located in the TOP 50/49 largest urban areas. Interestingly, these same respondents reported more frequent free exchange of data and less concern for cost recovery than the respondents from the random SAMPLE 64. They also displayed a slightly higher propensity toward coordinating their database activities.

Interorganizational GIS activities were guided by a variety of mechanisms, with some organizations often reporting multiple mechanisms. Technical groups, memoranda of understanding (MOU) or intergovernmental agreements (IGA), and various coordinating entities were used by about one-half of the respondents. But, generally, only for 24 respondents from the group of random SAMPLE 64 organizations and for 15 respondents from the TOP 50/49 largest urban areas (37.5 percent and 30.6 percent, respectively), the interactions were formalized with ordinances or resolutions of governing bodies, MOUs or IGAs, policies of elected officials or CEO/CAOs, and service or other contracts.

Establishment of rules and responsibilities was another indicator of the readiness to pursue joint data activities. The survey revealed only sporadic institution of joint rules and responsibilities for GIS database development, update, and use by local governments in the region. The results suggested that 45 respondents (70.3 percent) of the random SAMPLE 64 and 40 respondents (81.6 percent) from the organizations located in the TOP 50/49 sample of largest urban areas had no clearly assigned rules and responsibilities.

Next, funding was examined as the important operational aspect of implementing GIS technology and supporting data development, integration, and access. Three most frequently suggested sources of funding regional GIS included renewable grant funding, continuous financing (i.e., budget-line item), and contributions by local governments as part of regular programming/services, totalling about two-thirds of the funding sources. There was a noticeable difference between the funding sources for the random SAMPLE 64 respondents and the organizations from the TOP 50/49 largest urban areas. The latter had a substantial number of respondent organizations receiving continuous financing and renewable grants; the main source of funding for the former were the contributions by local governments.

Finally, the existence of state mandates for the development of land-information systems and the requirement for GIS use was explored. Seventeen or one-quarter of the random SAMPLE 64 respondents stated that there was a state land-information system program in place along with other related programs (14). Only a few respondents found the program funding to be adequate or somewhat adequate. GIS as a tool for developing land-information systems was required only in three cases. Fourteen programs in cadastral areas and 11 in other areas were reported by the organizations from the TOP 50/49 urban areas. The pattern was similar to the one that emerged in the SAMPLE 64—only a few considered the funding adequate and GIS was required only in one case.

People
The respondent organizations ranged in their professional staff size from a few to 140 for the random SAMPLE 64 and 450 for the TOP 50/49 group; the sample means were 18.7 and 52.5, respectively. The means dropped to the respective means of 3 and 4.6 professional full-time equivalent staff members working on GIS database development, maintenance, and integration.

Inquiry into the history of regional cooperation revealed that respondents from the TOP 50/49 largest urban areas on average had less positive experiences. A total of 17 (24.7 percent) of those organizations responded as having either an “excellent” or a “very good” history of cooperation; 22 (34.4 percent) of the respondents from the random SAMPLE 64 checked those two options. This result is not too surprising given that cooperation is more needed but also more difficult to realize in large metropolitan regions with more governmental entities and substantive and administrative complexities.

Communications were considered one aspect of regional cooperation. Many respondent organizations were part of a regional network that facilitated interaction between their members. In assessing their network’s functionality, about two-thirds of the respondents perceived it as a viable means of communication (i.e., strongly agreed, agreed, or were neutral). Still, about one-quarter of the respondents either “disagreed” or “strongly disagreed” about

---

*The percentages do not add up to 100 percent because respondents were asked to report multiple methods.
the network’s functionality. Overall, the respondents from the TOP 50/49 group were slightly more positive about the value of the network than the respondents from the random SAMPLE 64.

Interestingly, in general, higher engagement in cooperative projects was reported by organizations from the TOP 50/49 urban areas. Similarly, leadership support and commitment for cooperation and GIS data sharing seemed to be associated with those same respondents. Figure 5 summarizes the responses on the following statements from both the random SAMPLE 64 respondents and from the TOP 50/49 group of largest urban areas:

- In general, governments in your region often undertake joint cooperative projects;
- The political and government leaders in the region are committed to cooperation; and
- The government leaders in your region have been very supportive of geographic data sharing.

Exploring the Significance of Factors

Regression analysis was used to explore factors that affected the integration of data sets across metropolitan areas. All well-established and proven factors were included and significantly related to the number of layers available for the region as a whole (see Table 4).

Staffing support and provision of continuous funding were associated with a higher number of geographic information layers. The other stimulating factors, such as GIS mandates and pressing urban and regional issues, also played an important role in database integration at the regional level. As expected, the technological element in terms of ease of conversion between different data formats was not relevant. However, surprisingly, clearly defined rules and responsibilities were statistically significant but negatively related to the number of regionally assembled layers.

Finally, while the regional wealth did not seem to be associated with the establishment of an integrated database, the number of administrative entities (e.g., counties, parishes, boroughs) that a metropolitan area consisted of was crucial. The complexity of regional relationships increased with the number of entities involved and with it the challenges they faced.

<table>
<thead>
<tr>
<th>Dependent Variable: Number of GIS Layers Across the Jurisdiction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(Const)</td>
</tr>
<tr>
<td>Format conversion difficulty</td>
</tr>
<tr>
<td>Staff</td>
</tr>
<tr>
<td>Continuous financing</td>
</tr>
<tr>
<td>Rules and responsibilities</td>
</tr>
</tbody>
</table>

Table 4. Regression analysis of regional database integration (N = 64)

SUMMARY AND CONCLUSIONS

Information on spatial data holdings and regional assembly was solicited via a Web-based survey of 388 metropolitan planning organizations (MPOs) and other regional entities to assess regional GIS capacity. In addition to understanding the availability of relevant data themes in the metropolitan constituent units (counties, parishes, or boroughs), the extent of data assembly at the regional level, technological, people, policy, and contextual conditions for development and integration of regional GIS data sets were assessed. Summary and analysis of responses were based on two samples: 64 randomly received responses representing a cross section of metropolitan areas of different sizes and 49 responses from organizations located in the largest 50 urban areas.

The study results suggest that despite the major advancements in GIS technology and the extensive efforts spent in spatial data development at the local, regional, state, and national levels in the past two decades, the status of the regional GIS capacity may not be taking full advantage of what is technologically feasible. To begin with, regional data sets were available, assembled, and regularly updated in only a small segment (one-third or less) of metropolitan areas. There was a high compatibility in software and data formats and general openness to geographic data sharing, but limited use of advanced methods for data exchange and integration. For example, Internet-based access to data and establishment of clearinghouses still were rare. Also underdeveloped were the formalized interorganizational mechanisms and agreements on standards, rules, and responsibilities, with one-third or more respondents reporting the absence of such agreements. This finding is consistent with Harvey and Tulloch’s (2006) recent research on local data sharing, where they report a relatively low level of formalization of those relationships and activities.

The most interesting and unexpected result in the regression model was the negative significance of rules and responsibilities as a factor in favor of more extensive database assembly. Possibly, the fact that very few respondents had defined rules and responsibilities may have influenced this result. Also, this finding points to differences in sampling between this and the previous research by Nedović-Budić and Pinto. Namely, the sample in the previous research drew from the organizations that already were engaged in...
GIS coordination and interorganizational GIS, while this research drew on the population of all regional agencies, regardless of their GIS coordination activity. Perhaps in the organizations already coordinating their GIS, rules and responsibilities were commonly established and reinforcing (but not determining) the success of interorganizational activities; in the random sample where very few organizational groups had established rules and responsibilities, they turned out to be irrelevant for the outcome of joint database activities.

In addition to the establishment of standards and rules and responsibilities as the most concrete manifestation of the readiness to pursue joint data activities, regional cooperation and communication were considered fundamental for building regional GIS capacity. However, they as well were not fully functional in the majority of metropolitan environments. Other factors appeared as expected, such as the presence of regional problems related to urban sprawl and natural resources that stimulated the development and integration of regional GIS databases. Similarly, mandated programmatic requirements for GIS also were encouraging, although not adequately funded. Finally, the lack of stable funding for regional GIS was a general problem. A majority of regional database endeavors drew on local contributions and grants; continuous funding based on budget-line items was rare.

The organizations located in the TOP 50/49 largest urban areas functioned under slightly different circumstances than did organizations in medium and small urban areas. Their environments on average could be characterized as more affluent (and thus having more stable funding to support regional GIS efforts), more active in data development activity, more open to data sharing, and more likely to employ advanced methods (e.g., the Internet) for providing data access. However, the larger size of those regions and the higher number of participants involved posed substantial difficulties in cooperating, establishing functional communication networks, accessing other organizations’ data, and agreeing on mutual standards and roles and responsibilities. This finding is consistent with previous research on multiparticipant GIS, which recognizes that the increased number of participants complicates the GIS interactions (Ventura 1995). It also relates to the already proven account that coordination incurs cost (Kumar and van Dissel 1996) that grows with the size of the organizations involved.

This study provides one of the rare comprehensive national assessments of regional GIS capacity. It challenges the perception that GIS and spatial data are everywhere and ubiquitously available and accessible even in a country well advanced in the application of geospatial technologies such as the United States. The progress is slower than expected or perceived. Even though the picture probably has changed somewhat since the time the evidence presented in this paper was collected, the situation would not be radically different. Previous studies on the diffusion of GIS show that institutional change lags behind the technology (Budić and Godschalk 1994, Campbell and Masser 1995, Masser and Onsrud 1993). In addition, the regional GIS explored in this study requires interorganizational coordination activities that themselves present many challenges, including, among others, the balancing of diverse interests and motivations; technological, human, and financial resources and capacities; rules and responsibilities; as well as the particulars of each context (Nedović-Budić and Pinto 2000, 2001; Nedović-Budić et al. 2004a). While we are optimistic that the picture has improved since the data for this study was collected in 2003 (particularly with respect to Internet-based access to regional datasets), we also believe that the difference would not be dramatic.

The study findings also raise the question whether these regional GIS could qualify for SDIs and whether they present a substantial installed base for building the higher levels of SDI, e.g., the U.S. NSDI. To a large extent, the answer to this question depends on the definition of SDI. If it is equated with the presence of a clearinghouse, than many of the regional GIS explored here still are unqualified for SDI status; if there is a more relaxed understanding of SDI as a network of data producers and users who exchange spatial data in many ways, than regional GIS may be considered as SDIs and as building blocks of the NSDI. This conjecture along with the other research findings applies primarily to building of the U.S. NSDI, because of its unique institutional setting and culture. The latter is increasingly recognized as an important factor in understanding the nature of SDI developments around the world (Atlantic Institute 2005) and one of the key variables in any international comparative study and project (Hofstede 1980).

As predicted, NSDIs are not easy to develop and we have a long way to go (Masser 2005, Nedović-Budić and Pinto 1999b). The question of NSDI and its installed base is particularly intriguing and is only related to regional GIS but not explicitly addressed in this study. However, we suggest that local and regional GIS have merits and require attention as the constituents of such an installed base. In the United States, a consistent inclusion of this level of spatial data production and use is missing. Despite the FGDC’s attempts to reach beyond the state level, these efforts are not sufficiently systematic and binding as they are for the federal agencies. The partnership approach is valuable but limited in the environment that is competitive and fiscally tight. Progress toward the U.S. NSDI requires implementation of more creative and perhaps more aggressive programs geared toward the local and regional installed base. While funded mandates are proven mechanisms of motivating local and regional actions, there may be other innovative approaches to developing and sustaining a viable NSDI, such as restructuring or realignment of institutions dealing with spatial data, inventions in the legal realm, or pragmatic and productive financing schemes. Among other recent efforts is the formation of the National Geospatial Advisory Committee (NGAC) in 2008. The purpose of the group is to facilitate discussions about the framework for a national geospatial policy—outcomes as well as organizational structure (http://www.fgdc.gov/ngac/).

Finally, unlike Rajabifard et al. (2003) who consider the local and regional levels as primarily product-driven, we find that it is also interorganizational processes and interactions that give shape to the regional GIS (and consequently the regional SDI). The relationship between different levels of the NSDI hierarchy and the need for a bottom-up building of the NSDI should be explored in more detail. The topics to tackle include, but are not limited to: definition and
characteristics of local and regional SDI; connectivity and compatibility among SDIs at different levels; exploration and comparison of SDI hierarchical and network models; programmatic, legal, and financial support for enhancing local and regional capacity; relationship between interorganizational GIS and SDI; a link between the NSDI and its installed base; and innovations in institutionalizing and sustaining (spatial) information infrastructures. Ultimately and most importantly, it is the translation of the existing and future SDI research results into effective practice that matters the most.

About the Authors

Zorica Nedović-Budić is Professor Chair of spatial planning and GIS in the School of Geography, Planning and Environmental Policy at the University College Dublin.

Gerrit-Jan Knaap is Professor of Urban Studies and Planning and Executive Director of the The National Center for Smart Growth Research and Education at the University of Maryland.

Nama Raj Budhathoki is a PhD candidate in the Department of Urban and Regional Planning at the University of Illinois at Urbana-Champaign. His research focuses on volunteered geographic information (VGI).

Branko Cavrić is Associate Professor of city planning and GIS in the Department of Architecture and Planning, Faculty of Engineering and Technology at the University of Botswana.

Acknowledgments

The authors would like to thank the Lincoln Institute of Land Policy, the U.S. Department of Housing and Urban Development, and the Federal Highway Administration for their support for this project and their recognition that sound urban and regional planning is possible only by maintaining current and comprehensive GIS data. We also thank the National Association of Regional Councils and the Association of Metropolitan Planning Organizations for their advice and technical assistance. Our competent and committed graduate research assistants—Kate Crowley and Tolga T. Yilmaz at the University of Illinois and Elisa Vitale at the University of Maryland—provided valuable help with the project.

Last, but not the least, we could not do without the cooperation and generous input from the respondents to our Web survey. Their time and insight is greatly appreciated. Any opinions, findings, or conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the institutions involved.

References


National Geospatial Intelligence Agency (NGIA). Http://www.nga.mil/portal/nga01/.


### Appendix A

**SAMPLE DATA**

<table>
<thead>
<tr>
<th>DATA THEME</th>
<th>DATA AVAILABILITY</th>
<th>IS THE DATA SET ASSEMBLED?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td><strong>LAND</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land Use—Existing</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td>Land Use—Future or Planned</td>
<td>29</td>
<td>45</td>
</tr>
<tr>
<td>Parks and Public Lands</td>
<td>37</td>
<td>58</td>
</tr>
<tr>
<td>Open Space and Protected</td>
<td>27</td>
<td>42</td>
</tr>
<tr>
<td>Developed Land</td>
<td>26</td>
<td>41</td>
</tr>
<tr>
<td>Vacant Land</td>
<td>25</td>
<td>39</td>
</tr>
<tr>
<td>Infill Sites</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Brownfields</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Historic/Archaeological Sites</td>
<td>17</td>
<td>27</td>
</tr>
<tr>
<td>Tax Assessment Data</td>
<td>26</td>
<td>41</td>
</tr>
<tr>
<td><strong>REGULATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoning</td>
<td>20</td>
<td>31</td>
</tr>
<tr>
<td>Local Comprehensive Plan Designations</td>
<td>26</td>
<td>41</td>
</tr>
<tr>
<td>Building Permits (Georeferenced)</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Subdivision Approvals</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>Development Rights (Georeferenced)</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Growth Boundaries/Service Areas</td>
<td>24</td>
<td>38</td>
</tr>
<tr>
<td><strong>BOUNDARIES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Municipal</td>
<td>55</td>
<td>86</td>
</tr>
<tr>
<td>Annexation (Georeferenced)</td>
<td>27</td>
<td>42</td>
</tr>
<tr>
<td>School Districts</td>
<td>36</td>
<td>56</td>
</tr>
<tr>
<td>Special Districts</td>
<td>20</td>
<td>31</td>
</tr>
<tr>
<td>Parcel Boundaries</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td><strong>NATURAL FEATURES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rivers/Streams</td>
<td>50</td>
<td>78</td>
</tr>
<tr>
<td>Floodplains</td>
<td>36</td>
<td>56</td>
</tr>
<tr>
<td>Wetlands</td>
<td>32</td>
<td>50</td>
</tr>
<tr>
<td>Soils</td>
<td>31</td>
<td>48</td>
</tr>
<tr>
<td>Land Cover/Vegetation</td>
<td>28</td>
<td>44</td>
</tr>
<tr>
<td>Biodiversity/Habitats</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>Topography</td>
<td>35</td>
<td>55</td>
</tr>
<tr>
<td><strong>TRANSPORTATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Street/Road Network</td>
<td>53</td>
<td>83</td>
</tr>
<tr>
<td>Transit Lines</td>
<td>42</td>
<td>66</td>
</tr>
<tr>
<td>Ports (Air, inland, sea)</td>
<td>38</td>
<td>59</td>
</tr>
<tr>
<td>Traffic Analysis Zones (TAZ)</td>
<td>50</td>
<td>78</td>
</tr>
<tr>
<td><strong>UTILITIES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sewer Infrastructure</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>Storm Drainage Inf.</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>SERVICES</td>
<td>All</td>
<td>None</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>Number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schools</td>
<td>31</td>
<td>48</td>
</tr>
<tr>
<td>Schools Service Areas</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td>Fire Stations</td>
<td>24</td>
<td>38</td>
</tr>
<tr>
<td>Fire Service Areas</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td>Police Stations</td>
<td>23</td>
<td>36</td>
</tr>
<tr>
<td>Police Service Areas</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>OTHER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satellite Imagery</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>Aerial Photography</td>
<td>45</td>
<td>70</td>
</tr>
<tr>
<td>Real Estate Market Inf. (MLS)</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

**TOP 50/49**

<table>
<thead>
<tr>
<th>DATA THEME</th>
<th>DATA AVAILABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAND</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Land Use—Existing</td>
<td>26 53 3 6</td>
</tr>
<tr>
<td>Land Use—Future or Planned</td>
<td>20 41 9 18</td>
</tr>
<tr>
<td>Parks and Public Lands</td>
<td>30 61 2 4</td>
</tr>
<tr>
<td>Open Space and Protected Lands</td>
<td>23 47 4 8</td>
</tr>
<tr>
<td>Developed Land</td>
<td>18 37 7 14</td>
</tr>
<tr>
<td>Vacant Land</td>
<td>16 33 8 16</td>
</tr>
<tr>
<td>Infill Sites</td>
<td>5 10 21 43</td>
</tr>
<tr>
<td>Brownfields</td>
<td>2 4 22 45</td>
</tr>
<tr>
<td>Historic/Archaeological Sites</td>
<td>13 27 11 22</td>
</tr>
<tr>
<td>Tax Assessment Data</td>
<td>13 27 15 31</td>
</tr>
<tr>
<td>REGULATION</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Zoning</td>
<td>9 18 9 18</td>
</tr>
<tr>
<td>Local Comprehensive Plan Designations</td>
<td>13 27 13 27</td>
</tr>
<tr>
<td>Building Permits (Georeferenced)</td>
<td>8 16 19 39</td>
</tr>
<tr>
<td>Subdivision Approvals</td>
<td>6 12 18 37</td>
</tr>
<tr>
<td>Development Rights (Georeferenced)</td>
<td>2 4 27 55</td>
</tr>
<tr>
<td>Growth Boundaries/Service Areas</td>
<td>14 29 19 39</td>
</tr>
<tr>
<td>BOUNDARIES</td>
<td>All</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Municipal</td>
<td>41</td>
</tr>
<tr>
<td>Annexation (Georeferenced)</td>
<td>16</td>
</tr>
<tr>
<td>School Districts</td>
<td>26</td>
</tr>
<tr>
<td>Special Districts</td>
<td>12</td>
</tr>
<tr>
<td>Parcel Boundaries</td>
<td>14</td>
</tr>
<tr>
<td>NATURAL FEATURES</td>
<td>All</td>
</tr>
<tr>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Rivers/Streams</td>
<td>42</td>
</tr>
<tr>
<td>Floodplains</td>
<td>30</td>
</tr>
<tr>
<td>Wetlands</td>
<td>26</td>
</tr>
<tr>
<td>Soils</td>
<td>27</td>
</tr>
<tr>
<td>Land Cover/Vegetation</td>
<td>24</td>
</tr>
<tr>
<td>Biodiversity/Habitats</td>
<td>14</td>
</tr>
<tr>
<td>Topography</td>
<td>23</td>
</tr>
<tr>
<td>TRANSPORTATION</td>
<td>All</td>
</tr>
<tr>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Street/Road Network</td>
<td>39</td>
</tr>
<tr>
<td>Transit Lines</td>
<td>39</td>
</tr>
<tr>
<td>Ports (Air, inland, sea)</td>
<td>28</td>
</tr>
<tr>
<td>Traffic Analysis Zones (TAZ)</td>
<td>44</td>
</tr>
<tr>
<td>UTILITIES</td>
<td>All</td>
</tr>
<tr>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Sewer Infrastructure</td>
<td>4</td>
</tr>
<tr>
<td>Storm Drainage Inf.</td>
<td>4</td>
</tr>
<tr>
<td>Potable Water Inf.</td>
<td>5</td>
</tr>
<tr>
<td>Other Utilities</td>
<td>4</td>
</tr>
<tr>
<td>SERVICES</td>
<td>All</td>
</tr>
<tr>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Schools</td>
<td>29</td>
</tr>
<tr>
<td>Schools Service Areas</td>
<td>12</td>
</tr>
<tr>
<td>Fire Stations</td>
<td>18</td>
</tr>
<tr>
<td>Fire Service Areas</td>
<td>9</td>
</tr>
<tr>
<td>Police Stations</td>
<td>17</td>
</tr>
<tr>
<td>Police Service Areas</td>
<td>9</td>
</tr>
<tr>
<td>OTHER</td>
<td>All</td>
</tr>
<tr>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Satellite Imagery</td>
<td>16</td>
</tr>
<tr>
<td>Aerial Photography</td>
<td>35</td>
</tr>
<tr>
<td>Real Estate Market Inf. (MLS)</td>
<td>4</td>
</tr>
</tbody>
</table>
Appendix B

Availability and Update of Land-use and Street/Road Data (SAMPLE 64)

Availability and Update of Land-use and Street/Road Data (TOP 50/49)
Availability and Update of Aerial Photography and Satellite Imagery (SAMPLE 64)

Availability and Update of Aerial Photography and Satellite Imagery (TOP 50/49)
Hierarchical Clustering of Multiobjective Optimization Results to Inform Land-use Decision Making

C. M. Moulton, S. A. Roberts, and P. H. Calamai

Abstract: Many, if not all, land-use planning design problems involve multiple conflicting objectives. For these multiobjective problems, a single plan can be chosen by giving weights or priorities to the objectives, comparing pairs of plans, or generating all plans for which no absolutely better plan exists, i.e., the Pareto optimal set of plans. The advantage of the Pareto optimization approach is it delays judgments regarding the relative importance of the objectives until their interactions and trade-offs have been explored. Unfortunately, the Pareto optimal set may contain a very large number of plans. Choosing a plan for implementation or further consideration from such a large set is potentially intractable for a decision maker. Previous approaches to this problem in Pareto optimization aimed to find a representative subset of this solution set. The proposed methodology provides a tractable representation of Pareto optimization results for the land-use application. Hierarchical cluster analysis is used to organize the solutions into a tree structure based on their objective function values. This approach is applied to a land-use change problem in an urban fringe area in southern Ontario, Canada. An example decision demonstrates the use of the resulting structure in decision making. Applicability of this methodology for other applications and in a decision support system is discussed.

INTRODUCTION

Planning problems often have many potential solutions and multiple competing objectives. These types of problems are well addressed by multiobjective optimization methods. Multiobjective optimization is applied to problems in a variety of fields where multiple conflicting objectives must be considered. The result is a nondominated set of potential solutions. A solution is nondominated if no feasible solution exists that is better on all objectives.

Balling (2004) used a multiobjective optimization algorithm to consider city and regional land-use and transportation planning. Like this paper and that of Roberts (2003), the goal of using multiobjective optimization was to improve on traditional planning methods. In most planning decisions, the alternative plans are formulated based on the experience and preferences of planners then presented to the public and the decision makers. This small set of plans cannot adequately capture the complexity of the planning problem and is inherently subjective. To evaluate their approach, Balling (2004) presented the results of their analysis to local city planners, state planners, and environmental planners, as well as local politicians. All persons consulted approved of this approach and encouraged continued work although a final plan was not chosen from the 100 resulting plans. Motivating this work, Balling (2004) believes that one reason that a plan was not chosen from the optimization results is the difficulty of considering such a large number of plans. Even with a large number of plans for consideration, planners uncovered key aspects of the problem that were used in the selection of a final plan. With current computing power, it is possible to consider multiobjective problems with very large nondominated solution sets. For example, the land-use problem in this paper has 6,561 nondominated solutions. According to Balling (2004), the number of plans to be considered must be objectively reduced to a set of plans representing “distinct conceptual ideas.” In other words, decision makers need a sample of the nondominated solutions that is sufficiently representative of the possibilities and trade-offs but small enough for tractable consideration.

The multiobjective optimization literature acknowledges the need for a method of reducing or organizing the nondominated set (Benson and Sayin 1997). Several researchers (Rosenman and Gero 1985, Morse 1980, Taboada et al. 2007) have dealt with this issue using cluster analysis or filtering. This paper differs from their work for it aims not only to make the nondominated set tractable but to do so without removing any elements of the nondominated set before presenting the solutions to the decision makers. A set of potential land-use plans is organized into nested groups of similarly performing plans without filtering out any plans.

Cluster analysis can be applied to the results of a multiobjective optimization algorithm to organize or partition solutions based on their objective function values. In this paper, clustering is used to take a large set of land-use plans and organize them based on proportions of urban, natural, and agricultural land-use as well as landscape-ecology measures. The goal of clustering is to create an “efficient representation that characterizes the population being sampled” (Jain and Dubes 1988, p. 55). Such a representation allows a decision maker to further understand the decision by making available the attainable limits for each objective, key decisions and their consequences, and the most relevant variables; this presentation is an improvement on a list of potential solutions and their associated objective function values.

This paper details a hierarchical cluster analysis approach to organize Pareto optimization results into a hierarchical representation. The methodology is applied to a greenlands system design problem in an urban fringe area in southern Ontario, Canada. In this study,
171 unique potential land-use configurations were generated using the Nondominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al. 2002) described in the following section with eight landscape ecology--based criteria. A hierarchical algorithm is desirable for it presents a nested organization of the land-use configurations that can be used for decision making as demonstrated in the example decision.

**MULTIOBJECTIVE OPTIMIZATION**

A multiobjective framework is used for optimization problems that cannot be satisfactorily converted into a single objective problem. Within this framework, trade-offs are explicitly considered. A multiobjective optimization problem is composed of a set of decision variables, a set of objective functions of those variables to be maximized or minimized, and a set of constraints on the values of those variables.

Without loss of generality, assume that all objective functions are to be maximized. Mathematically, a multiobjective problem can be written as shown in problem 1.

Maximize \( f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})) \)

Subject to \( \mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{X} \)

where \( \mathbf{X} \) is the set of feasible solutions, often defined by a set of constraints.

It is unlikely that a single solution \( \mathbf{x} \) in \( \mathbf{X} \) maximizes all the objective functions simultaneously because the objective functions typically conflict. The efficient set, \( \mathbf{E} \), is the set of feasible solutions \( \mathbf{x} \) in \( \mathbf{X} \) for which no other feasible solution is as good as \( \mathbf{x} \) with respect to all objective functions and strictly better than \( \mathbf{x} \) in at least one objective function. For \( I = \{1, \ldots, m\} \), the efficient set is formally defined as in equation 2.

\[
\begin{align*}
\mathbf{E} = & \{ \mathbf{x} \in \mathbf{X}: f_i(\mathbf{x}) \geq f_i(\mathbf{y}) & \forall \mathbf{y} \in \mathbf{X}, i \in I, \text{ and} \\
& f_j(\mathbf{x}) > f_j(\mathbf{y}) \text{ for some } i \in I \}
\end{align*}
\]

The solutions in \( \mathbf{E} \) are said to be Pareto optimal or globally nondominated. The Pareto front is the mapping of the efficient set to the space defined by the objective functions, i.e., \( \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{E} \} \). A nondominated set is a set that is efficient with respect to its own elements, i.e., satisfying equation 2 with \( \mathbf{E} = \mathbf{X} \). No solution in a nondominated set dominates or is dominated by any other solution in the set. In Figure 1, solutions \( A \) and \( B \) are nondominated for no solutions are better on both objectives \( f_1(\mathbf{x}) \) and \( f_2(\mathbf{x}) \). Solution \( C \) is dominated by solution \( A \) but not by solution \( B \), because \( C \) is better on objective \( f_1(\mathbf{x}) \) but \( B \) is better on objective \( f_2(\mathbf{x}) \). The efficient or nondominated set consists of solutions \( A \) and \( B \).

**SOLUTION METHODS FOR MULTIOBJECTIVE OPTIMIZATION**

Three approaches can be taken to find a solution to multiobjective problems (Benson and Sayin 1997). The first approach entails reformulating the problem as a single objective problem. To do so, additional information is required from the decision makers, such as the relative importance or weights of the objectives, goal levels for the objectives, value functions, etc. The second approach requires that the decision makers interact with the optimization procedure typically by specifying preferences between pairs of presented solutions. The third approach, Pareto optimization, finds a representative set of nondominated solutions approximating the Pareto front. Pareto optimization methods, such as evolutionary multiobjective optimization algorithms, allow decision makers to investigate the potential solutions without a priori judgments regarding the relative importance of objective functions. Post-Pareto analysis is necessary to select a single solution for implementation (Benson and Sayin 1997).

All three approaches to solving multiobjective optimization problems have shortcomings. The solution returned by the single objective approach can highly depend on the weights and, in non-convex problems, the responses to changes in weights or goals may be unpredictable. With conflicting and noncommensurate criteria as well, it can be difficult to make value judgments such as choosing weights or goals for the criteria. Given decision maker input, the first approach returns a single solution. Interactive approaches consider only a small set of nondominated solutions because of the effort required (Benson and Sayin 1997). Pareto optimization approaches return a potentially large number of solutions for consideration. Selecting a single solution from a large nondominated set is likely to be difficult for any decision maker. Benson and Sayin (1997) proposed that an ideal solution procedure for multiobjective optimization is to provide the decision makers with a globally representative subset of the nondominated set that is sufficiently small so it is tractable. This work aims to approach this ideal procedure by accepting the computational effort required to generate a large nondominated set and subsequently organizing it based on its structure in such a way that decision makers can tractably consider interesting subsets without a priori removal of any solutions from consideration.

**Figure 1. Example of dominated and nondominated solutions**

- \( f_1(\mathbf{x}) \)
- \( f_2(\mathbf{x}) \)
- A
- B
- C

\( A \) and \( B \) are nondominated for no solutions are better on both objectives \( f_1(\mathbf{x}) \) and \( f_2(\mathbf{x}) \). Solution \( C \) is dominated by solution \( A \) but not by solution \( B \), because \( C \) is better on objective \( f_1(\mathbf{x}) \) but \( B \) is better on objective \( f_2(\mathbf{x}) \). The efficient or nondominated set consists of solutions \( A \) and \( B \).
NSGA-II ALGORITHM

Any Pareto optimization method could be employed in our methodology. NSGA-II is used for it is known to perform well with nonconvex, disconnected, and nonuniform Pareto fronts (Deb et al. 2002). The use of this heuristic algorithm allows for efficient searching through many potential solutions while considering several objectives that may be nonconvex and discontinuous.

NSGA-II is an evolutionary multiobjective algorithm, specifically a genetic algorithm. Evolutionary multiobjective algorithms apply biologically inspired evolutionary processes as heuristics to generate nondominated sets of solutions. It should be noted that the solutions returned by evolutionary multiobjective algorithms may not be Pareto optimal, that is, globally nondominated, but the algorithms are designed to evolve solutions that approach the Pareto front and spread out to capture the diversity existing on the Pareto front to obtain a good approximation of the Pareto front.

Genetic algorithms operate on a population of solutions and employ selection, crossover, and mutation operators, among others, to generate successive improved populations based on a fitness function. At each generation, a set of potential parents is generated, subsets of the parents are combined to create offspring, and the fittest offspring are included in the next generation (Falkenauer 1998).

NSGA-II differs from single objective genetic algorithms in two respects: It aims to maintain diversity in the population instead of converging to a single solution and it uses nondomination to assess the fitness of individuals. NSGA-II also is elitist; the members of the next generation can be drawn from either the offspring or the parents of the current generation, thus potentially preserving good solutions from the previous generations that may have been destroyed by crossover or mutation.

The procedure for applying NSGA-II and an example follow:
1. Randomly generate the first generation.
2. Partition the parents and offspring of the current generation into k fronts $F_1, F_2, \ldots, F_k$ so that the members of each front are dominated by all members of better fronts and by no members of worse fronts.
3. Calculate the crowding distance for each potential solution: For each front, sort the members of the front according to each objective function from the lowest value to the highest value. Compute the difference between the solution and the closest lesser solution and the closest greater solution, ignoring the other objective functions. Repeat for each objective function. The crowding distance for that solution is the average difference over all the objective functions. Note that the extreme solutions for each objective function are always included.
4. Select the next generation based on nondomination and diversity:
   - Add the best fronts $F_1, F_2, \ldots, F_j$ to the next generation until it reaches the target population size.
   - If part of a front, $F_{j+1}$, must be added to the generation to reach the target population size, sort the members from the least crowded to most crowded. Then add as many members as are necessary, starting with the least crowded.
5. Generate offspring:
   - Apply binary tournament selection by randomly choosing two solutions and including the higher-ranked solution with a fixed probability, typically between 0.5 and 1 (Goldberg and Deb 1991).
   - Apply single-point crossover and the sitewise mutation to generate offspring.
6. Repeat steps 2 through 5 for the desired number of generations.

An example of the Pareto ranking and crowding distance calculations for a two-objective function maximization problem is shown in Figure 2. In this example, the population size is six so there are 12 solutions (parents and offspring) considered. Six of these solutions are included in the next generation. First, the Pareto fronts are identified: The first front contains the nondominated solutions, the second front contains the solutions dominated only by the first front, and the third front contains the solutions dominated by the first and second fronts. The next generation is formed by taking the three solutions in the first front and the three least-crowded solutions in the second front. These solutions then would be used to generate offspring and the process would be repeated.

POST-PARETO ANALYSIS

Post-Pareto analysis aids decision makers in choosing a single solution from the potentially large set of Pareto optimization results. Several researchers have applied clustering methods in different ways to nondominated sets to aid decision makers. Most of these methods use the similarity of elements in the nondominated set based on their objective function values and remove elements that are too similar to other elements. We construct a tree data structure for the nondominated set, allowing decision makers to
consider subsets of the nondominated solutions prior to removal of any solutions from further consideration.

Mattson et al. (2004) detailed a "smart Pareto filter" to obtain a representative subset of a nondominated set by defining regions of "practically insignificant trade-offs" around solutions. Each solution is considered successively and all solutions with "practically insignificant trade-offs" are removed for they are not sufficiently distinguishable from the solution under consideration. To create a representative set, more elements of the nondominated set are retained in areas with steeper trade-offs, commonly known as "knees." The dimensions of the regions of "practically insignificant trade-offs" for each objective function must be specified a priori (Mattson et al. 2004).

Morse (1980) detailed one of the first applications of cluster analysis to a nondominated set. The multiobjective programs considered were linear programs. A solution was removed from the nondominated set if it was indistinguishable from another solution based on decision maker–defined thresholds. Morse (1980) evaluated seven hierarchical clustering methods. Ward's method, the group average method, and the centroid method performed very well. Ward's method was preferred because the clusters at the same level of the hierarchy were similar in size and shape, although it performed only slightly better than the centroid and group average methods (Rosenman and Gero 1985). Ward's method creates clusters by minimizing the variance within each of the two clusters at each level of the hierarchy.

Rosenman and Gero (1985) applied complete linkage hierarchical clustering to "reduce the size of the Pareto optimal set whilst retaining its shape" (p. 189). This method allowed control of the diameter of the resulting clusters. They noted that solutions whose vectors of objective function values are similar may have decision variable vectors that are similar or very different but this idea was not further explored. The objective functions were considered successively to avoid the implicit aggregation in applying proximity measures. First, elements of the nondominated set were clustered using a single criterion. If a solution within a cluster dominated another solution in the cluster on all criteria except the clustering criterion, then the dominated solution was eliminated from consideration. The process was repeated for each criterion until the nondominated set was sufficiently small.

Taboada et al. (2007) used partitional (k-means) clustering for combinatorial multiobjective problems. Either the most interesting cluster, i.e., the "knee" cluster, was considered in detail by discarding the solutions in other clusters, or one solution from each of the k clusters was considered to form a representative subset of the nondominated set.

This paper differs from the previously mentioned work by considering hierarchical clustering and not reducing the size of the nondominated set under consideration. The hierarchical tree structure for the solutions allows the decision makers to tractably consider the solutions using a sequence of decisions to reduce the set of solutions under consideration.

**CLUSTER ANALYSIS**

Cluster analysis involves using algorithms and techniques to examine the internal organization in a data set in an objective way; it can be used to describe the data concisely and to uncover patterns and relationships that may not be readily apparent (Dubes 1993). The aim is to group objects that are similar in some way. A cluster analysis requires several steps: data scaling, proximity calculation, selection of a clustering algorithm, application of the clustering algorithm, and validation and consideration of results.

The input data is represented as a matrix, X, containing the p criteria, e.g., the objective function values, of the n elements to be clustered, e.g., the Pareto optimization results. Without data normalization or scaling, the relative values of the objective functions may act as implicit weightings. This weighting is undesirable because Pareto optimization is used to generate an unbiased set of optimal trade-off solutions without considering the relative importance of the objective functions. Milligan and Cooper (1988) and Gnanadesikan (1995) found range scaling better for recovering known cluster structures than raw data or normalization to unit variance and zero mean. Range scaling was used in our proposed hierarchical clustering methodology.

Clustering methods can be separated into two categories: partitional methods that provide a single partition of the solutions and hierarchical methods that provide a series of nested partitions. With a hierarchical clustering method, the number of clusters need not be known a priori (Ward 1963).

The tree structure of a hierarchical clustering algorithm can be useful for guiding decision processes when many alternatives must be considered. The tree of the cluster hierarchy often is represented in a dendrogram, where the top element in the tree, the root, is a cluster containing all the elements and the bottom elements, i.e., leaves, represent individual elements. The dendrogram displays the merging (or dividing) of clusters from the leaves to the root (or the root to the leaves) and the distance or dissimilarity between the merged (or split) clusters. The dendrogram is an objective structure that can be used by decision makers to discuss and consider the clustered elements.

The group average linkage method with Euclidean distance was used for this analysis. The Euclidean distance is easily interpretable and invariant to rotations and translations (Dubes and Jain 1976). The group average linkage computes the distance between clusters as the mean distance between all pairs with one element in the first cluster and one element in the second cluster including duplicate elements. Hierarchical clustering linkage methods, like all clustering methods, make assumptions about the sizes and shapes of clusters (Jain et al. 1999). Different linkage methods tend to identify clusters with different characteristics. The group average linkage allows clusters to vary in size and shape.

Validation, beginning with establishing a clustering tendency in the input data, is important because clustering methods will find clusters even in random data (Dubes and Jain 1979). The Cluster Validation and Analysis section that follows includes the process of validating the clustering results.
**METHODOLOGY**

The proposed hierarchical clustering methodology is summarized in this section. The steps of this analysis are as follows:

1. Define decision variables, feasible set, and objective functions.
2. Choose and apply a Pareto optimization algorithm, e.g., NSGA-II.
3. Cluster analysis:
   - Clustering tendency: By visual inspection or data projections verify that a hierarchical cluster structure is a reasonable model for the data.
   - Data scaling: Remove implicit variable weightings due to relative scales using range scaling.
   - Proximity: Select and apply an appropriate similarity measure for the data, here, Euclidean distance.
   - Choice of algorithm(s): Consider the assumptions and characteristics of clustering algorithms and select the most suitable algorithm for the application, here, group average linkage.
   - Application of algorithm: Apply the selected algorithm and obtain a dendrogram.
   - Validation: Examine the results based on application subject matter knowledge, assess the fit to the input data and stability of the cluster structure, and compare the results of multiple algorithms, if used.
4. Represent and use the clusters and structure: If the clustering is reasonable and valid, examine the divisions in the hierarchy for trade-offs and other information to aid decision making.

The greenlands design problem is described in the following section, followed by the results of the cluster analysis and an example decision.

**GREENLANDS DESIGN PROBLEM**

The greenlands design problem detailed by Roberts (2003) is applied to an urban fringe area near Toronto, Ontario, Canada. In this region, single-family residential housing and aggregate extraction (hereafter referred to collectively as urban), agriculture, and natural areas coexist. The analysis aims to inform land-use decision making concerning the effects of land-use, in particular potential habitat loss and fragmentation represented by reduction in the area and connectedness of natural land. The model takes into account the existing landscape features and land-use (see Figure 3). Currently abandoned fields could potentially be used for urban growth, reseeded or allowed to regenerate as natural areas, or restored as agricultural land; these fields are the candidate sites for land-use change. The configuration of the landscape features is important for the function of the landscape for natural systems and human use.

The problem is formulated as a configuration optimization problem. Configuration optimization is a class of combinatorial optimization that manipulates geometric and topological properties of a system to optimize the system performance (Roberts 2003). Categorical variables represent the land-use classes. Twenty-eight total land-use classes in the source data were aggregated into four classes denoted as {1, 2, 3, 4}, representing “unchanged,” “natural,” “agriculture,” and “urban.” There are eight candidate sites available for land-use change as shown in white on the study area plot in Figure 3.

Eight objective functions were formulated based on landscape-ecology metrics (Roberts 2003). A nonmathematical description of each objective function follows. Many of the objective function evaluations depend on graph data structures representing the topology of the study area. Each objective function is formulated for maximization.

**GA1 Area of Natural Features:** More natural area is better. This objective is implemented as the ratio of the area of the candidate sites coded “natural” and the total area of the candidate sites.

**GA1 Area Weighted Mean Shape:** Compact natural areas are more desirable than elongated natural areas. This principle is modeled by maximizing the mean area to the perimeter ratio of the n largest sets of connected natural polygons. In this study, n is five, although n only reaches four in this data set.

**GA2 Natural Feature Connectivity:** Connected natural sites are preferable to the same natural sites scattered across the landscape. This objective maximizes the mean number of connected natural sites in the n largest connected sets of natural sites.

**GA3 Stepping-stones of Natural Features on Shortest Paths:** Paths of natural sites through the landscape allow for flora and fauna mobility. The number of natural sites along n(n-1)/2 “stepping-stone” shortest paths between the n largest natural areas is maximized.

**GA4 Patches of Natural Features within Urban Areas:** Patches of natural area within urban areas are desirable. This objective maximized the number of links between urban sites and natural sites within urban areas based on spatial autocorrelation join counts.
GA5 Agricultural Area: The area of the candidate areas assigned to agriculture is maximized. This objective is implemented as the ratio of the area of the candidate sites coded “agricultural” and the total area of the candidate sites.

GA6 Clustered Development: More compact urban areas are more desirable. Similar to objective GA4, this objective maximizes the number of urban to urban adjacencies and is implemented based on join counts.

GA7 Urban Area: Similar to objectives GA1a and GA5, this objective competes for land-use. It is implemented as the ratio of the area of the candidate sites coded “urban” and the total area of the candidate sites.

Because the decision variables are categorical, the true Pareto front is a discrete set of solutions. Given this discreteness and the nonlinearity of some of the objective functions, the density of the solutions may not be homogeneous across the front. Because of this variation in solution density, a hierarchical clustering structure may exist.

RESULTS
This section describes the application of the hierarchical clustering methodology presented previously to the greenlands design problem using NSGA-II. To assess the quality of the NSGA-II results, a full enumeration of the Pareto front was obtained. This full enumeration was possible because of the small sample area considered.

The NSGA-II results contained 171 unique solutions from the four to the eight (4^8 or 65,536) different possible land-use configurations from eight candidate sites and four potential land-use categories. Figure 4(a) shows the NSGA-II results and Figure 4(b) shows the objective function values for the 6,561 solutions on the fully enumerated Pareto front. The fully enumerated Pareto front was obtained by removing the dominated solutions from the potential solutions. Ideally, the NSGA-II results would have the same range and distribution as the values in the true Pareto front. As seen in Figures 4(a) and 4(b), the ranges and distributions of the approximation and the true values are very similar. The differences between the ranges and distributions of these values are discussed below. Note that objective GA4, patches of natural features within urban areas, takes only a single value in these results because of the small problem size and the existing land-use and is excluded from further analysis.

NSGA-II is a heuristic method to approximate the full Pareto front. It trades off time and the number of solutions, allowing the consideration of larger problems, with precision. The NSGA-II approximation of the Pareto front contains approximately 2 percent as many solutions as the true Pareto front (171 versus 6,561).

There are several discrepancies between the approximation of the Pareto front shown in Figure 4(a) and the true Pareto front shown in Figure 4(b). The true Pareto front must contain the solutions that would result from optimizing each of the objective functions separately while ignoring the other objectives. In this problem, the percentage land-use area objective functions, GA1a, GA5, and GA7, would each attain a value of one. These solutions were not included in the NSGA-II results. However, it is unlikely that these extreme solutions are politically acceptable, as per the example decision that follows. There also were proportionally fewer solutions in the upper portions of the ranges for the natural and agricultural land-use objective functions, GA1a and GA5, respectively. The mean value for the urban land-use objective function, GA7, was higher in the NSGA-II results than in the enumeration of the true Pareto front. NSGA-II explored the span of potential values for the land-use objective functions, but there is a bias toward urban land-use.

NSGA-II can generate a good approximation of the Pareto front with less time and computational effort. It provides a small sample covering the large set of potential solutions. This ability allows considering larger problems than would otherwise be feasible.

CLUSTER VALIDATION AND ANALYSIS
The first step of the cluster analysis is establishing the existence of a clustering tendency. In Figure 5, each objective function is plotted against each other objective function. Evidence of clusters is clear in several of the plots. For example, considering GA1a plotted against GA5, three large clusters are apparent: one cluster with low values of GA1a and GA5, one cluster with high values of GA1a and low values of GA5, and one cluster with low values of GA1a and high value of GA1a. These same three major clusters also can be seen in the plots of GA1a against GA7 and of GA5 against GA7. The attribution of land to the differing land-uses is an important characteristic of this decision and the presence of these major clusters should be detected by any successful clustering algorithm. Several subclusters can be seen within each of the major clusters, confirming the expected hierarchical cluster structure. For example, in the cluster where both GA1a and GA5 take low values, there are five well-separated dense regions. At this point in the methodology, an indication of a clustering tendency is established. Without an indication of a clustering tendency, the results of applying a clustering algorithm may not reflect any pattern in
the data. The true structure may not correspond directly to the obvious clusters in these plots because of relationships between multiple objective functions that are not easily visualized. For example, all three of the land-use percentage objective functions must be considered to understand the trade-off for the percentage of land allocated to each land-use. The proposed clustering methodology considers these simultaneous interactions between multiple objective functions.

Next, the scale of the data was addressed. All but one of the objective functions range from 0 to 1. Only objective function GA6, measuring clustered development using spatial autocorrelation, could take values beyond this range. GA6 was rescaled to lie in the range 0 to 1 by linearly mapping the values in the NSGA-II results to the range 0 to 1.

The Euclidean distance was used as the measure of similarity between vectors of objective functions. For this application, the weighted group average method was found to be the most suitable clustering method. This method was applied to the NSGA-II results, giving the dendrogram shown in Figure 6. Beginning at the root, each split of the dendrogram into two subclusters can be qualified in terms of the differences between the subclusters, for example, C(2,1,1) for a cluster derived by choosing the second cluster at the first branching, the first cluster at the second branching, and the first cluster at the third branching.

The cophenetic correlation coefficient compares the distance between solutions and their relative position in the dendrogram. In this case, its value was 0.9247, indicating a good fit of the data to the dendrogram. Stability testing further validated the resulting clustering structure. Three types of stability tests were applied:

- Adding random error terms of up to 25 percent,
- Omitting up to 25 percent of the solutions, and
- Randomly splitting the data into two sets equal-sized before clustering.

The first and second branchings shown in the example decision were robust to these perturbations. Although in some of the more extreme perturbations, the order of these branchings was reversed with a split between natural and agricultural land occurring before the split between natural and urban land.

Three features are important to the success of this cluster analysis: The cluster structure must be a valid representation of the data, obvious clusters must be detected, and, where no obvious clustered exist, it must segment the clusters by reflecting the structure of the data.

This methodology succeeded in identifying the obvious clusters. Cluster C(1) contained the solutions with high values of objective function GA7, urban land-use area, which only take low values of GA1a, natural land-use area, and low to moderate values of GA5, agricultural land-use area. Branching cluster C(2) results in a trade-off between GA1a and GA5, the natural and agricultural land-use area objective functions. C(2,1) had low values of GA1a and high values of GA7 while C(2,2) had high values of GA1a and low values of GA7. The three major clusters were identified in the first two branchings as C(1), C(2,1), and C(2,2). In cluster C(2,2), there was no obvious branching into two subclusters. The clustering algorithm branched the cluster into two subclusters so that the solutions in cluster C(2,2,1) were preferable to those in cluster C(2,2,2) on objective function GA6, clustered development on which they all attained the maximal value. As well, no solution in cluster C(2,2,1) took the minimal value for objective function GA5, agricultural land-use area. The solutions in cluster C(2,2,2) attained similar or better values of objective function GA1a, natural land-use area, than the solutions in cluster C(2,2,1), and similar or worse values of objective function GA5, agricultural land-use area. None of the solutions in cluster C(2,2,2) took a value of zero for objective function GA7, urban land-use area, and no solution in cluster C(2,2,1) included any new urban land-use area. These subclusters were clearly different and reflected trade-offs between the objective functions.

Internal, external, and relative validity were considered. The internal validity of the weighted group average linkage method results was satisfactory. The cophenetic correlation coefficient of 0.9247 was sufficiently large to indicate a good fit of the dendrogram to the data. Error perturbation, data deletion, and data split stability tests indicated that the three major clusters that were detected are

---

**Figure 5.** Scatter plots of objective function values

**Figure 6.** Weighted group average linkage method dendrogram
a valid structure. The order of the first two branchings giving these three clusters was less robust as reflected in the similar heights in the original dendrogram. The clusters caused by branchings resulting lower in the dendrogram were less robust than the clusters from higher branchings. This reduced robustness reflected that those clusters are less differentiated and in many cases they were not the result of an obvious cluster structure.

The external validity was established by considering the clusters in relation to knowledge about the landscape configuration problem. The cluster analysis results reflected the real-world aspects of the landscape configuration problem, indicating good external validity. The clusters of landscape designs were differentiated in terms of the land-use codings for the candidate sites. An example of this assessment can be found in the example decision that follows.

The relative validity was assessed by comparing the hierarchical weighted average linkage clustering results to a similar method, hierarchical complete linkage clustering. The weighted group average linkage method results tended to agree with the complete linkage method, indicating good relative validity. The discrepancies in the results from these methods can be attributed to known assumptions of the complete linkage method, in particular similarly sized clusters that did not agree with the visual inspection of the cluster structure.

These assessments confirmed that there were three major clusters in the data and that the general dendrogram structure was valid although the clusters lower in the dendrogram were less robust. The remainder of this section details an example decision process based on the results of the cluster analysis.

**EXAMPLE DECISION**

This section provides an example of the use of the hierarchical clustering structure in the land-use decision problem. The local human population requires land to work, live, and grow food, i.e., urban and agricultural land. Land at the fringes of the currently developed areas is desirable for these purposes, but this land currently serves natural functions. For example, water recharge areas and animal habitats may exist within the natural area. In this paper and previous work (Roberts 2003), the natural functions that require specific land parcels, such as water recharge, were dealt with by preprocessing, and the natural functions that may not be specific to particular land parcels, such as some animal habitats, are dealt with using the multiojective optimization model for the landscape configuration problem. Known specific desired habitat areas also could be preprocessed as natural in this methodology. In this scenario, there are several current candidate sites whose land-use can be changed and the most significant concern of the decision makers is the loss of natural land related to the first four objective functions and GA6, which measures the clustering of urban development. The clustering of urban development is also desirable for human use of the land, for example, services such as public transit and waste collection can be implemented more efficiently in compact urban areas.

The example decision begins by considering the branching into clusters C(1) and C(2), and then proceeds to the preferred cluster and considers that branching. This process is repeated until there is a small set of landscape designs amenable for further, more detailed consideration.

**FIRST BRANCHING**

*Observations:* Figure 7 shows the objective function values of the solutions in the two clusters resulting from the first branching at the root of the dendrogram. The trade-off in land area for the different land-uses is evident: Cluster C(1) contains the solutions with high values of objective function GA7, urban land-use area, which only coincide with solutions with low values of GA1a, natural land-use area, and low to moderate values of GA5, agricultural land-use area. The solutions in C(1) achieve a wide spread of values for objective function GA6, ranging from approximately 0 to 0.8. Cluster C(2) contains the solutions with low values of GA7, urban land-use area. Cluster C(2) does not restrict the values of objective functions GA1a, natural land-use area, and GA5, agricultural land-use area. Similar to cluster C(1), the solutions in cluster C(2) take a wide range of values for GA6, clustered development, for more configurations are available with more sites coded urban, but in cluster C(2) the values for GA6 range from approximately 0.2 to 1. Note that in this particular study area, none of the candidate sites are adjacent to existing urban areas (see Figure 3) so any new urban area will be less compact.

*Decision:* The algorithm extracted the expected conflict between natural and urban land-use. C(1) is dominated by urban development while C(2) contains solutions with various levels of agricultural and natural land. The consequences of favoring urban or nonurban land can be seen in the differences between the two clusters on the other objective functions. In this particular problem, favoring natural or urban land-use has little impact on the natural landscape-ecology objectives (GA1, GA2, and GA3). Because the more natural plans on cluster C(2) may result in highly clustered
urban areas, in this example we further consider cluster C(2). The cluster or clusters for further consideration can be chosen with the knowledge of the relationships between the objectives for this particular land-use problem. It should be noted that for other problems, the natural-urban division may not occur at the first branching; for example, the first branching could divide clusters on GA1 if candidate sites could connect the existing natural areas. In this problem it implies that this conflict is the dominant effect in this decision.

**SECOND BRANCHING**

*Observations:* The result is a trade-off between GA1a and GA5, the natural and agricultural land-use area objective functions, respectively (see Figure 8). C(2,1) has low values of GA1a and higher values of GA5, while C(2,2) has high values of GA1a and lower values of GA5.

*Decision:* Again this example problem demonstrates expected conflicts: in this case, agricultural versus natural land-use. One might thus choose cluster C(2,2) for further consideration for the natural land-use is the highest priority in this decision scenario. In contrast to selecting weights for the objectives or singularly emphasizing new natural area, there are a number of plans to consider for which the trade-offs between objectives have been made explicit. The effects of emphasizing natural area on urban clustering and the natural landscape functions were available for consideration before the decision makers needed to narrow the set of plans under consideration.

**THIRD BRANCHING**

*Observations:* Figure 9 shows the clusters resulting from branching cluster C(2,2). The solutions in cluster C(2,2,1) are preferable on objective function GA6, clustered development, on which they all attain the maximal value. Also, no solution in cluster C(2,2,1) takes the minimal value for objective function GA5, agricultural land-use area. The solutions in cluster C(2,2,2) attain equivalent or better values of objective function GA1a, natural land-use area, and equivalent or worse values of objective function GA5, agricultural land-use area. Some of the solutions in cluster C(2,2,2) do not take a value of zero for objective function GA7, urban land-use area, but no solution in cluster C(2,2,1) includes any new urban land-use area.

*Decision:* Further consider cluster C(2,2,1) in this scenario with the understanding that none of the new land is allocated for urban use. Consider also choosing C(2,2,2) for further consideration for it has a small quantity of new urban land while noting that the amount of new agricultural land may be reduced and that allowing any new urban land will degrade the clustering of the urban development.

At this point, the solutions may be deemed sufficiently similar in performance that other factors that were unmodeled by this methodology, for instance land costs and availability, aesthetics of viewsheds, etc., must be considered. The insight from considering the trade-offs in the objective functions can be included in the decision, but the details of the landscape designs should be considered. Cluster C(2,2,1) contains only two landscape configurations and cluster C(2,2,2) contains ten landscape configurations. Figures 10 and 11 show the plots for the solutions in clusters C(2,2,1) and C(2,2,2).

In agreement with the emphasis on natural land-use, the largest candidate site, site four, is natural in all of these plans. Within cluster C(2,2,1), sites three and six also are always natural and site five is always agricultural. Within cluster C(2,2,2), site one is unchanged or agricultural and at least one of the small sites is urban. While the solutions in both of these clusters are very similar, the superior performance of cluster C(2,2,1) on the clustered development objective...
function corresponds to the lack of new urban land. Within the clusters, the land-use of the larger sites is consistent and the plans are mostly differentiated on the land-use of the smaller sites. For objective function GA1, the area weighted shape of natural area, none of the solutions in clusters C(2,2,1) or C(2,2,2) take the lowest values attained for this objective function; in all of these solutions having site four as natural land improves the shape of the largest natural area. Within cluster C(2,2,1) and for solutions A, C, F, H, and I in cluster C(2,2,2), the smaller natural area above the center of the study area has an improved area weighted shape because of the natural land-use of site six. In plan F in cluster C(2,2,2), the natural area weighted shape for the largest natural area is improved by having site five as natural. Within clusters C(2,2,1) and C(2,2,2), the natural area stepping-stone shortest paths measured by objective function GA3 always outperforms the worst attainable value. Like the natural area weighted shape, this improvement is because of the additional natural areas.

In the first few branchings of the dendrogram, the clusters correspond to those noted in the visual inspection for clustering tendency. The branchings lower in the dendrogram do not correspond to visually obvious clusters for no obvious clusters exist. These branchings segment the obvious clusters into subclusters that are differentiated but not significantly separated. The use of a hierarchical linkage clustering algorithm allows the method to deal with these branchings where there may be no cluster structure and return usable results.

**DISCUSSION**

There are three primary aims for the proposed methodology. First, it should create a tractable presentation of the NSGA-II results for the landscape configuration problem; visually obvious clusters should be detected and a useful structure should be provided even where no obvious structure exists. The validity of the resulting structure also is important. This aim was discussed in the previous section. Second, it should be adaptable to other problems suited to a multiobjective optimization framework using Pareto front enumeration or approximation methods. This second requirement includes being extendable to include other model aspects such as constraints, preferences, and weights. Third, it should be simple enough to be understood by decision makers, including the general public, and to be potentially included in future decision support systems without extensive training. If the system is a black box, it will not be accepted and used in the expected decision making contexts.

Using the binary branching structure in the dendrogram, the solutions can be considered based on their objective function values. Potentially interesting subsets of solutions for further consideration can be found by reducing the set under consideration by descending in the tree from the root until a sufficiently small set of solutions with sufficiently similar objective function values remains. Using
the dendrogram resulting from the weighted group average linkage method, the set under consideration can be made arbitrarily small. Because the tree is not balanced, the decrease in the number of solutions under consideration resulting from each branching is not predictable and many branchings may need to be taken to obtain a sufficiently small set. If desired, a different linkage method, such as the complete linkage method, could be employed to return a more balanced dendrogram that is less indicative of the trade-off surface structure. In the landscape configuration problem considered here, three branchings typically are sufficient to reduce the solutions under consideration to a small compact set. At that point, unmodeled aspects of the decision, such as the suitability of individual candidate sites for more specific uses, should be considered.

The proposed methodology provides a tractable representation of the multiobjective optimization results. While the effects of the relative ranges of the objective functions may complicate the use of this methodology or implicitly convey additional importance to a particular objective function, currently there is no available method for considering multiple objective functions simultaneously without some consideration of the relative importance of the objective functions. Here the emphasis is on those objective functions that clearly differentiate the clusters occurring higher in the dendrogram.

CONDITIONS FOR REUSE AND EXTENSION
The characteristics of the input data are important to the use and success of this methodology. The proposed methodology is most easily applied where the input data is a Pareto front derived from an approximation algorithm or enumeration and a clustering tendency is seen in two-dimensional visualizations. If the decision variables in the multiobjective optimization problem are continuous, the input to the clustering algorithm must be a discrete approximation of the Pareto front.

If the problem has only two or three objective functions, and in particular if those functions are well behaved, i.e., convex and continuous, there is little benefit to using the proposed methodology. If a simple two-dimensional or three-dimensional visualization of the Pareto front or a good approximation thereof is available, the proposed methodology cannot lead to additional insight. The proposed methodology is particularly useful where there are more than three objective functions but not so many objective functions that it becomes difficult to select one of the clusters at a branching.

If the decision variables are continuous or discrete with a constant density and a good approximation of the Pareto front is easily obtained, then the approach taken by Matsson et al. (2004) to find “interesting” regions of the Pareto front may be more suitable. Applying the methodology proposed in this paper still may yield insight for these problems, particularly if it is difficult to obtain a good approximation of the Pareto front. If no clustering tendency exists, then any structure resulting from the application of a clustering algorithm will be an artifact of that clustering algorithm. Nonetheless, a clustering algorithm could be used to construct a tractable representation using a dendrogram.

SUITABILITY FOR DECISION SUPPORT SYSTEMS
Once a hierarchical clustering structure is obtained, the dendrogram can be used in decision support systems. The dendrogram may be enhanced by simultaneous display with other visualizations to aid in enabling insight into the problem of interest. Seo and Shneiderman (2002) present uses for dendrograms in exploring high-dimensional hierarchical cluster structures in the context of genomic microarray analysis. One visualization using a dendrogram is to display the dendrogram and use columns of color blocks below each leaf to display information relevant to that leaf. In the landscape configuration example, this visualization could be used to display the land-use codings for each solution from NSGA-II. Figure 12 shows the solutions in cluster C(2,2,2), employing this visualization using colors yellow for agricultural, green for natural, gray for urban, and white for unchanged. The dendrogram provides an order for the solutions that allows the differences and similarities in the land-use codings to be seen relative to similarities in the objective functions represented by the dendrogram. Using this enhanced dendrogram would give insight into key sites contributing to objective functions and allow the user to verify that aspects of the problem are properly modeled.

For example, in the landscape configuration problem, the connectivity of the core natural areas is important. A small number of sites may determine this feature of the landscape design. Using a dendrogram enhanced with a color block view of the candidate site land-use codings would allow users to see whether particular sites tend to be similar within clusters and different between clusters. For deployment in a decision support system, further work would be required to design appropriate visualizations.

A possible different use of the dendrogram for visualization in the landscape configuration problem is to use it as an input interface to allow users to display the full maps of the study area. Choosing a cluster would overlay the land-use codings in the NSGA-II solution, allowing the user to see the solutions of interest as a whole landscape design.

This methodology is expected to be used in an iterative decision process, where the problem is reformulated based on the output of earlier iterations. Objective functions and constraints...
on the decision variables can be added, removed, or changed and the analysis repeated. This iterative process ensures that the model accurately represents the problem and explores the problem to obtain additional insight. An iterative process can be used to allocate limited resources to investigating potential solutions. At the first iteration, the proposed methodology is applied to a small sample of the feasible solutions. The proposed methodology can be applied to each of these interesting regions in turn by constraining the decision variables or placing limits on the objective function values. Computational effort need not be expended on exploring the entire solution space in detail for the shape of the trade-off surface can be considered. While this methodology could be implemented using off-the-shelf tools such as SPSS, R, and S, the effort to do so could not be easily justified unless the methodology were key to the decision process. Commercial development could allow for considering larger problems as well as improving usability through a modern graphical interface, including real-time interactive visualizations of solutions. If real-time interaction is not necessary, this methodology can be scaled up to include 20,000 or more partitions (Roberts 2003). The future work detailed in the following section discusses these issues to be addressed prior to easily generalizing the methodology for other objectives and problems.

CONCLUSIONS AND FUTURE WORK

Pareto optimization methods allow using multiobjective optimization models without a priori decision maker preferences. The decision makers can consider the possibilities and trade-offs between objectives before selecting a solution for implementation. These methods suffer from the shortcoming of requiring the decision makers to consider many possible solutions resulting from the optimization procedure. This paper developed and evaluated a cluster analysis methodology to address this issue. A land-use planning problem was used as motivation and to evaluate the proposed methodology.

Previous work in multiobjective optimization in land-use planning called for a method to objectively determine a set of plans representing “distinct conceptual ideas” (Balling 2004). Previous methods involved eliminating some of the Pareto optimal solutions before presenting them to the decision makers. The proposed methodology allows the entire nondominated set to be presented to the decision makers by providing a tractable structure for the results. This methodology will continue to be applicable as computational power increases and Pareto optimization algorithms improve, allowing the consideration of larger nondominated sets.

This approach is applicable to multiobjective problems with discrete decision variables or hierarchically clustered nondominated sets. Multiobjective configuration optimization problems and the more general class of combinatorial multiobjective optimization problems have discrete Pareto fronts. It also may be applicable to problems containing highly discontinuous Pareto fronts. If a hierarchical structure is not suspected in the data or if the structure is not to be used in the decision process but a clustering tendency exists in the data, then the methodology presented by Taboada et al. (2007) involving k-means clustering to create a partitioning of the solutions into a predefined number of clusters may be more suitable.

The proposed methodology is particularly useful if similarly performing solutions based on the objective function values may be distinguishable to the decision makers based on the importance of the decision variable values or unmodeled aspects of the problem. Previous approaches to this issue would have eliminated similarly performing solutions from consideration.

Future work will revisit the issues in cluster analysis, including scaling, proximity measures, selection of algorithms, and validation, as well as improved visualizations. The limitations of the small study area and the structure resulting from the particular landscape ecology metrics used also require further consideration. This work could be extended to consider the proximity of the solutions based on their decision variable values, e.g., in the land-use application the similarity of the landscape configurations. It may be desirable in some applications to highlight clusters containing similarly performing solutions with very different decision variable values; these solutions could denote unmodeled aspects of the problem or possible freedom in the decision. Shape space measures (Small 1996) may be a suitable approach to comparing the morphology of the landscape configurations.

About the Authors

C. M. Moulton received her Master’s degree in Applied Science in Systems Design Engineering at the University of Waterloo. Her research interests include geomatics, data mining, decision support, and multiobjective optimization.

Steven Roberts is an associate professor in the Department of Geography and Environmental Studies at Wilfrid Laurier University where he teaches geomatics and spatial analysis courses at the graduate and undergraduate level. His Ph.D. is in Systems Design Engineering from the University of Waterloo and he has undergraduate degrees in Urban and Regional Planning (Waterloo) and Mathematics (McMaster). His current research interests include spatial data models and data structures, combinatorial optimization, genetic algorithms and programming (Evolutionary Multicriterion Optimization), applied graph theory and category theory, landscape ecology, and parallel computing.

Corresponding Address:
Wilfrid Laurier University
Department of Geography and Environmental Studies
Room 3C11, Arts Building
Waterloo ON Canada N2L 3C5
Phone: (519) 884-1970, Ex. 2470
Sroberts@wlu.ca
www.wlu.ca/-wwwgeog/facstaff/SRoberts.html
P. H. Calamai is a professor in the Department of Systems Design Engineering at the University of Waterloo. He received his Ph.D. in Systems Design Engineering from the University of Waterloo. His research interests include facility location and resource allocation as well as multidisciplinary design optimization and decision support systems.

Acknowledgments

Support for this work was provided by the Ontario Graduate Scholarship and the Department of Systems Design Engineering at the University of Waterloo. The second author acknowledges GEOIDE grant #HSS-DSS-17 for partial support of this work.

References


The Potential of Integrating E-participation in Planning Support Systems

Arjen Koekoek, R. van Lammeren and G. Vonk

Abstract: The increasing complexity of spatial planning issues and pressure from citizens to take part in deciding on spatial plans result in a need for improved methods to aid communication between governmental actors and citizens. These developments put high demands on planning support systems (PSS), instruments that can aid planners in performing their planning tasks. By using the accessibility of the Internet, e-participation seems an attractive shoot of PSS. This integration could facilitate citizen involvement in planning. Although many advantages are attributed to participatory PSS, its use in the planning practice remains marginal until now. According to this paper, this is partly caused by the lack of empirical studies that demonstrate potential benefits and obstacles when applying participatory PSS. To help alleviate this problem, this paper synthesizes three case studies of Dutch municipalities that each implemented some type of e-participation to achieve particular goals. It demonstrates that although e-participation has potential as a PSS, a limited impact on decision making and a feedback provision currently constitute obstacles for effective participation.

INTRODUCTION

Changing social and political conditions and the trend toward a democratization of environmental decision making make it necessary to reconsider the role of participation in planning (van den Brink et al. 2007). Citizens increasingly expect to see their voices reflected in decision making. A recent survey (Ernst and Young 2008) reveals that more than four of every five Dutch citizens would like to have a say in important decisions on the municipal level. Traditional nonparticipatory approaches to spatial planning appear to be insufficiently capable of creating the societal support necessary to implement plans, often causing resistance and delays. Since the 1990s, a “communicative turn” in planning can be observed, necessary to cope with the changing needs of society (Healey 1993). This trend toward more interactive and participatory planning is likely to have major repercussions on the way planning is practiced. Some authors expect that planning will become more complex and increasingly depend on information technology instruments (Brail 2008, Geertman 2002b, Geertman and Stillwell 2009). In our view, this seems in line with the demand to handle the information, views, and opinions from so many more voices in the efficient and effective way that we expect from authorities today.

The Web 2.0 trend pressures governments to open up their decision-making processes for citizens to participate over the Internet, in so-called e-participation. E-participation has the potential to establish more transparency in government by allowing citizens to use new channels of influence that reduce barriers to public participation in policy making (UN 2008). In concordance with others (Al-Kodmany 2003, van den Brink et al. 2007), participation is perceived here as a two-way interaction between government and the public. Advantages of e-participation tools over traditional participation tools are that communication no longer is bound to a specific location and a specific time. Tools for e-participation can be categorized in discussions and chats, polls, and (GIS-based) visualizations (Lenos and Buurman 2000). The use of visualizations, especially three-dimensional, is interesting for they are easier for common citizens to understand than are policy documents (Riedijk and Van de Velde 2006). According to Klos- terman (2001), the search for an appropriate role for (a GIS-based) computer-based information and methods in planning must not begin with a particular technology but rather with a conception of planning. Planning support systems (PSS) have been defined as a subset of geoinformation technologies, dedicated to support those involved in planning to explore, represent, analyze, visualize, predict, prescribe, design, implement, monitor, and discuss issues associated with the need to plan (Batty 1995). Because of the more participatory nature of planning practice, the demand for PSS is likely to change. To meet the demands of participation, PSS increasingly need to facilitate reasoning together, retrieve empirical information, work community support, and disseminate knowledge (Geertman 2006). These are all characteristics in which GIS-based e-participation, situated at the focal point of e-government, public participation GIS, and planning support systems, should excel.

However, various studies underline the limited use of PSS for participation in the planning practice (Geertman 2002a, Jankowski and Nyerges 2003, Laituri 2003b, Sieber 2006, Dunn 2007, Kingston 2007, Geertman and Stillwell 2009). How can this be explained? Some argue that citizens are not prepared to effectively participate because of digital and spatial literacy (Eshet-Alkalai 2004, Laituri 2003a). Barber (1997, p. 224) argues that the trouble with the zealots of technology as an instrument of democratic liberation is not their understanding of technology but their grasp of democracy. This statement also seems to apply to participatory PSS, as Geertman (2006) and Jankowski and Nyerges (2003) signal a supply-side bias in research. A change in the focus for participation research, therefore, is justified and needed, shifting the attention

1 For the sake of readability in the rest of the paper, the term e-participation will refer to electronically enabled (GIS-based) participation applications.
toward the users of participation—government and citizens—and their needs.

The reason for an authority to apply participation can be instrumental—using participation as a means to achieve a policy aim—as well as normative—participation as an aim in itself (De Graaf 2007). Woltjer (2002) makes a further distinction in functions: Participatory planning can contribute to efficiency and effectiveness because it yields information and ideas, and because it enlarges public support for the decision and thus averts implementation problems, objections, and appeal. Table 1 provides examples of the functions of participation. Some studies highlight the potential of e-participation to give citizens a say in decision making (Al-Kodmany 2003, Geertman 2002a), or utilize citizens’ local knowledge (Dunn der Eijk and Bos 2007), or prevent objections (Moody 2007). For e-participation, these functions have not yet been extensively evaluated from an academic perspective.

Table 1. Functions of participation

<table>
<thead>
<tr>
<th>NORMATIVE Function</th>
<th>Examples</th>
<th>INSTRUMENTAL Function</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functioning of democracy</td>
<td>Give citizens a say in decision making</td>
<td>Influence</td>
<td>Give citizens a say in decision making</td>
</tr>
<tr>
<td></td>
<td>Involve politically marginalized groups</td>
<td>Effectiveness</td>
<td>Utilize local knowledge</td>
</tr>
<tr>
<td></td>
<td>Inform citizens</td>
<td>Efficiency</td>
<td>Prevent objections</td>
</tr>
</tbody>
</table>

Source: Adapted from Woltjer (2002).

Experiences from real planning examples, therefore, are necessary to provide municipalities information on the potential of e-participation. This study will attempt to make a contribution to close the knowledge gap between the application and the process by identifying obstacles that block the use of e-participation as a PSS. First, a framework is presented that can be used to identify obstacles in the e-participation process. This framework is used to guide the case study research in which we focus on three Dutch municipalities that started experimenting with e-participation.

E-PARTICIPATION IN THE PLANNING PROCESS

The perspective of technology acceptance is useful to identify obstacles that can block the widespread use of a technology. Frambach and Schilleweart (2002) identified five chronological stages (awareness, consideration, intention, adoption decision, and continued use) that a technology has to pass to be accepted by an organization. Vonk et al. (2005) concluded from a survey among experts that for PSS the main bottlenecks in this adoption process consist of limited awareness among planners of the existence of PSS and the purposes for which it can be used; a lack of experience with PSS and its potential benefits; and a low intention to start using PSS among possible users. The study of Vonk et al. (2005) took PSS in general as a starting point for research. But e-participation as a specific type of PSS situates it within participatory research and planning and therefore the nature of participatory processes itself requires more attention (Craig et al. 2002). An e-participation application is identical to any other PSS for it has to go through the same five stages, but with the multitude of stakeholders involved in its use, the application significantly differs from nonparticipatory PSS, resulting in a number of additional obstacles associated with the use of e-participation in the planning process. Innovation adoption literature does not provide suggestions on how to investigate obstacles associated with the use in a participatory planning process. This paper, thus, takes the participatory process as a starting point to investigate obstacles.

Before we can identify obstacles blocking these functions of participation, it is useful to take a closer look at the position of e-participation as a PSS. The conceptualization of the role of e-participation in a planning process starts with a concept of participation itself. The four criteria for participation specified by Brezovsek (1995) are a starting point to define e-participation in the planning process. According to these criteria (1) individuals (citizens) should be included, (2) participation is voluntary, and (3) it should refer to a specific activity, which is (4) directed toward influencing the authorities.

Following these criteria, in a typical e-participation process as considered in this paper, a (local) authority attempts to include citizens in the process, some of these citizens decide to participate and do so using an e-participation application, resulting in citizen input that will affect decision making. Along with others, participation thus is perceived as a process that eventually should result in the exertion of influence on decision making (Craig 1998, Harris and Weiner 1998, Kingston 2007, van den Brink et al. 2007). It is presupposed here that the final decision-making abilities remain with the municipality, but the degree in which the citizen input reflects in this decision differs. The potential of e-participation as a PSS is fully utilized if municipalities successfully involve the targeted citizens; these citizens can effectively participate using the application and receive feedback on the way their input reflects in decision making. In the Netherlands, real success stories of such full utilization are scarce. In a wider context, real success stories also seem to be scarce, although some healthy partnerships between agencies and citizens have occurred (Geertman and Stillwell 2009). This brings up the question where obstacles occur in a participatory process. To get a grip on where these obstacles prevail in the participation process, a framework was constructed from literature, identifying four possible obstacles that can block effective participation in the planning process. The resulting conceptual framework is visualized in Figure 1. The following section introduces the obstacles associated with the identified stages in the planning process: involvement of public, possibilities to participate, impact on decision making, and feedback provision, and shortly reflects on the scientific debate around these topics.
OBSTACLE I: INVOLVEMENT OF PUBLIC

A necessary question to start with addresses the interest of the authority in what public exactly should be involved. Schlossberg and Schuford (2005) categorize two possible criteria: those affected by a decision or program or those who can bring important information to a decision or program. Either way, both definitions are directed toward particular groups, which logically implies that some others are excluded. Good governance implies that citizens, depending on their interest, may expect that they become involved. Sieber (2003) suggests that use of e-participation, by definition, succeeds when as many community members as possible can utilize spatial information in the public decision-making process. Tackling the question of what constitutes the public in e-participation becomes especially difficult with Web-based applications that are designed to expand public outreach (Sieber 2006). The anonymity of the Web blurs the identity of the citizens. To maintain a degree of control over the citizen input, municipalities can use different types of (local) media to stimulate citizens to use the applications. Additionally, when offering services online, developers need to take the impatient behavior of the user into account. Citizens seem unwilling and cautious to register or download programs (Moody 2007). Opposing viewpoints exist regarding the effect of e-participation on the normative function of participation. Some consider Internet access problems as the most important disadvantage of e-participation. Citizens without Internet access or with limited computer skills are excluded from participation, reducing the representative value of the citizen input (Mayer et al. 2005, Moody 2007, Obermeyer 1998). Others see online participation as an opportunity to involve groups that are underrepresented in traditional meetings (Carver et al. 2001, Kingston 2007). But what people are exactly underrepresented? A Dutch study of the Dutch Traffic Advisory Agency (AVV 2003) focusing on the reasons for people not to participate in traditional meetings revealed that motivations can be categorized into five groups. More than half of these nonparticipants do not have a problem with participation in itself but with the way participation takes place (see Table 2). If e-participation offers opportunities to participate at the time and place of choosing, and at one’s own pace, it has the potential to address traditional nonparticipants. Addressing these politically marginalized groups through e-participation could be considered a goal in itself.

<table>
<thead>
<tr>
<th>Types of Nonparticipants</th>
<th>Motive</th>
<th>Percentage (AVV 2003)</th>
<th>Involvement through E-participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distrustful</td>
<td>Do not believe in participation</td>
<td>34%</td>
<td>Not plausible</td>
</tr>
<tr>
<td>Busy</td>
<td>Do not have time to participate</td>
<td>27%</td>
<td>Plausible</td>
</tr>
<tr>
<td>Researchers</td>
<td>Need time to research plan backgrounds</td>
<td>18%</td>
<td>Plausible</td>
</tr>
<tr>
<td>Unsure</td>
<td>Feel unsure about their opinion</td>
<td>10%</td>
<td>Plausible</td>
</tr>
<tr>
<td>Indifferent</td>
<td>Do not care about participation</td>
<td>10%</td>
<td>Not plausible</td>
</tr>
</tbody>
</table>

OBSTACLE II: POSSIBILITIES TO PARTICIPATE

The second barrier consists of the empowerment potential. A supporting PSS instrument should assist and not hinder the user in the process of giving one’s opinion (Geertman 2002a, Jankowski and Nyerges 2003). If citizens decide to participate using the application, their input is determined by at least three factors. First of all, the possibilities for participation are limited by the technical aspects of the application. This means that the instruments should be at least transparent, understandable, and user-friendly for people to participate successfully (Geertman and Stillwell 2003). Secondly, the possibilities for participation are affected by functionality of the application. An often-used categorization of citizen participation levels is composed of three levels: nonparticipation, tokenism, and citizen power (Arnstein 1969). In terms of this categorization, not all applications enable the highest levels of citizen participation. The format of the application determines the way people can express themselves, for example, by voting in polls or starting discussions. But, secondly, the possibilities also can be limited by the political will to empower citizens. Studies on community development projects involving public participation highlight this relation, suggesting that cultural and political context rather than hardware and software are the main obstacles to successful public participation in decision making (Craig et al. 2002, van den Brink et al. 2007).
OBSTACLE III: IMPACT ON DECISION MAKING

The total amount of citizen response gathered via the application ultimately should find its way in the decision-making process. But participation in the creation of citizen input does not necessarily give any power to those involved in, and affected by, the decision making (Aitken and Michel 1995). This last step, therefore, might be the most crucial one in the process. Critics argue that use of the technology lends the illusion of control over decision making when actual control remains within the governing class (Sieber 2006). If the citizen input does not penetrate in the decision-making process or if the use of the citizen input is not communicated back to the citizens, the risk of backfire exists. In other words, as Carver et al. (2001) formulate: How do planning authorities ensure that information reaches local people and that genuine responses from local people are acted upon? Edelenbos (2005) suggests that there is a “missing institutional link” between the interactive process and the formal municipal decision-making process. For example, he concludes that in the Netherlands, interactive governance needs better institutional linking to prevent the interactive process from becoming meaningless and useless in formal decision making. Participation results now are often set aside in formal decision making for a range of reasons, which makes the participatory process look like window dressing.

OBSTACLE IV: FEEDBACK MECHANISM

The fourth obstacle originates from the third obstacle. For e-participation to be successful, governments should not merely allow citizens to voice their views online; it is more important to construct a feedback mechanism that shows citizens that their views are taken seriously (UN 2008). Citizens will judge an interactive process primarily by the degree of direct or indirect influence they are able to exercise (Mayer et al. 2005). Government thus should inform citizens about the way their input reflects in decision making. If this feedback link does not exist, the risk of cherry-picking may occur (Edelenbos 2005). Decision makers will pick a selection of citizen contributions and include these in the decisions. This can make the rich diversity of the total citizen input evaporate.

METHODOLOGY

Little is known about the importance of the identified obstacles in the e-participatory processes. For that reason, we select suitable e-participation case studies to find out what functions were originally intended and what obstacles were recognized that could block effective participation. In February and March of 2008, we conducted a quick scan on the Web sites of the 100 largest Dutch municipalities. Each municipal Web site was scanned for 20 minutes to find applications that enabled e-participation in a two-dimensional or three-dimensional environment. In this time span, different searching strategies were applied, both by performing queries in the municipal search engine, as well as by using the Web menu. Although many municipal Web sites use GIS technology, only seven municipalities used the technology in an interactive way, giving citizens the opportunity to discuss and suggest spatial designs. Four of these municipalities applied Virtuocity, two applied the application E-spraak, and one applied Second Life. For the case study research, one municipality was selected per application (see Figure 2). The three selected municipalities all have about 100,000 inhabitants. All three applications were intended to function as additional channels for participation, used parallel to a traditional more formal participation process. The developed framework offers the possibility to score the three case studies on functions of participation and evaluate the degree in which the obstacles prevail. In April and May of 2008, five involved professionals were interviewed. We held semistructured interviews with both the municipal process managers and the application developers. The interviews focused on both the functions of e-participation and the role of obstacles. This division into two topics also is used to present the results. Additionally, formal and informal documents concerning the cases were used. The following section introduces the three cases.

Virtual Helmond Helmond was the first of four municipalities to introduce a virtual city in 2006 (http://www.virtueel-helmond.nl). The city was involved in a national subsidy program and thus had a considerable budget for innovative information and communication technologies (ICT) projects. Out of four competitors, the Virtuocity application, developed by Cebra, was chosen. This application was selected for it would best match two important goals of Helmond. First of all, the city needed a way to communicate proposed changes for inner-city redevelopment with the inhabitants. These inhabitants typically had little education and were expected to have difficulties interpreting two-dimensional maps. An additional reason of the municipality for searching for a new tool was the frustration about the domination of traditional participation meetings by a vocal few. The application gives a three-dimensional design of the proposed spatial changes. Citizens can freely move around in this virtual world and can compare the old and new situations using panoramic photographs. Participation is enabled by discussion forums, chatting, and occasionally voting polls for the choice of designs. The project has been initiated by the municipality of Helmond. To log in, a citizen first has to download a plug-in and pick a character. The Web site still is online and is regularly refreshed when new designs are ready.
E-spraak Maastricht The municipality of Maastricht applied E-spraak (http://www.espraak.nl/Maastricht) as a first step to consult citizens for a new bicycle plan before starting the official planning procedure. The municipality required a participatory platform to receive citizens' suggestions for improved cycle-friendliness of the city. Because most cycling suggestions are geographically specific, a general forum discussion or survey was not likely to return the required input. The Web application E-spraak, developed by the company Goudappel Coffeng, was selected for this purpose. E-spraak is a two-dimensional application that enables citizens to start discussions on specific locations—for example, to signal dangerous crossings. Local discussions appear as thumbnails on the map, so other citizens can react. The municipality started using E-spraak because of the associated disadvantages of traditional participation meetings: the stereotype of the older, highly educated white male participant and meeting domination by a vocal few. In the end of 2007, during a month, citizens could give input for the cycling plan. To react, people had to register and leave their name, user name, and mailing address. No specific downloads were necessary to participate.

Second Life Zoetermeer To strengthen its image of an ICT-innovative city, the municipality of Zoetermeer searched for a new communication channel. Unlike the previous two municipalities, Zoetermeer used an already existing platform (http://www.secondlife.nl) as offered by Linden Lab. Citizen participation is not the sole purpose of the municipality with Second Life. City branding and attracting business are other equally important goals. Second Life is a virtual world with users worldwide. Because the application was not intended to enable citizen participation, the application developer was not interviewed in this case. Developers constructed a three-dimensional replica of the town hall in Zoetermeer. Zoetermeer officially opened this electronic town hall in March of 2007 as the first municipality worldwide. Before users can visit Zoetermeer in Second Life they need to install the program and register. The possibilities for participation are limited to attending virtual meetings where land-use plans are discussed.

RESULTS

Functions of E-participation

The e-participation applications facilitate different functions of participation. For Helmond, Maastricht as well as Zoetermeer, frustrations with the traditional methods for participation were an important reason to introduce e-participation. The municipality respondents were asked to rank the application on the functions of e-participation derived from literature in Table 1 (see Figure 3). The application E-spraak seems best suitable to utilize the local knowledge of citizens in the process and give citizens a say in decisions. Citizens know best which cycling situations in the city are unsafe and what other problems occur. Virtual Helmond seems more suitable to increase the involvement of citizens in policy and address marginal groups. In Second Life, the participation is limited to normative functions. An interesting result is that both E-spraak and Virtual Helmond prevent objection and appeal. Especially the more or less “objective” representation of the future situation in Virtual Helmond makes people less suspicious than do design sketches. Ironically, the city council of Helmond was initially reluctant about using the high degree of detail, thinking it could cause protests on every plan detail. The municipalities argue that informing citizens remains an important aspect of the application. Both Maastricht and Helmond claim that use of e-participation leads to better decision making. Zoetermeer does not claim this.

Reflecting on the Obstacles: the Opinion of Municipalities

How do the municipal process managers themselves observe the obstacles? Figure 4 gives an overview of perceived obstacles (italic). The officials do not seem to regard involving the citizens as an important obstacle. An important reason for this is that they use the e-participation process parallel to a traditional participation process. The interviewees also do not regard technical restrictions as a factor that limits possibilities for participation. However, a lack of political support to fully utilize the applications potential...
can be observed, especially in Zoetermeer and Helmond. In both cases, the applications offer possibilities for higher levels of participation, yet the municipalities choose not to utilize these. Also the reflection of citizen input in decision making is limited. All municipalities use the applications to obtain an indication of what the average person thinks, rather than as a basis to guide spatial changes. For these three cases, citizens received little feedback on the input they delivered.

**OBSTACLE 1: INVOLVEMENT OF PUBLIC**

The cities of Maastricht, Helmond, and Zoetermeer used multiple media to inform citizens about the possibility to e-participate. In all cases, the front page of the municipal Web site, the local newspaper, and press releases were used to involve citizens, and in Helmond local television also was used. In Maastricht, 322 people registered, resulting in more than 800 reactions. In Helmond, 30 to 40 people visited the virtual city per day, up to hundreds after updates; in Zoetermeer, around 20 per day visited. Downloading the needed software and registration efforts did not seem to discourage citizens for these rates are far higher than the number of citizens participating in traditional ways. All the municipalities tried to involve as many citizens as possible and did not object if citizens from other municipalities participated. This approach seems to fit in best with Siebers’ (2003) recommendation to involve as many community members as possible.

Although limited access to the Internet and little IT knowledge are believed to exclude large groups from participation, the municipality representatives argue that the traditional methods of participating seem to exclude an even larger group. More than half of the citizens probably will never attend a traditional participation meeting. Helmond made sure that people without Internet access or having difficulties with the application could visit an information center in the city center. Computers and assistance were made available there. When comparing the demographic characteristics of traditional participants and e-participation users, the available data indicates that the latter tend to be less dominated by older, highly educated males (see Table 3). Both Maastricht and Helmond consider this an important strength. The city of Maastricht even suggests that users of the application form a better representation of society than the participants in traditional meetings. Because of the limited availability of user characteristics and the rough character of the data in Table 3, it is not yet possible to justify this statement. These user statistics are not available for the Second Life application.

**OBSTACLE 2: POSSIBILITIES TO PARTICIPATE**

The input of the citizens is first restricted by the format of the application. In all cases, reactions are monitored and censured. In practice, this is hardly necessary. In E-spraak, citizens can put locations on the agenda and react on discussions started by other citizens. The municipality did not interfere in this process. Citizens have the possibility to vote to agree or disagree with reactions of others. Although available, this last function was not used by the municipality when the reactions were analyzed. In Virtual Helmond, the forum was hardly used by citizens. The reactions on the forum mainly consisted of questions, answered by the municipality. Some citizens used the opportunity to chat with the aldermen and walk with them through the virtual world to give their opinion or to ask questions. In one occasion, citizens could vote for the design of playgrounds, choosing from three types of designs. This city considered using DigiD² but eventually choose not to, because the city feared this would repel many people. Instead, the city chose to limit the amount of votes to two per IP address. Overall, citizens had little opportunity to actually have a say in decision making using the application. This had more to do with the political will than the functionality of the application. Tilburg, another city using the same Virtuocity application, decided to take participation a step further, allowing people to vote for the design of the main city square. In total, more than 4,000 people voted and the winning design will be constructed. As a municipality communication adviser put it, “Technically seen, participation can already go much further, the application offers this functionality, but the political will to do this in Helmond does not yet exist” (Interview, Van den Berg 2008). In Zoetermeer, the possibilities to participate are limited to virtually attending participation meetings. All three applications currently are still under development, increasing the opportunities to participate by giving citizens the option to add pictures (E-spraak), enabling citizens to build their own designs (Virtuocity), and enabling citizens to rank three-dimensional urban redevelopment projects (Second Life).

---

**Table 3. User characteristics in traditional participation and e-participation**

<table>
<thead>
<tr>
<th>Source</th>
<th>Traditional Participation Meeting</th>
<th>E-spraak (N = 737)</th>
<th>Virtueel Helmond (N = 53)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male %</td>
<td>75%</td>
<td>67%</td>
<td>40%</td>
</tr>
<tr>
<td>Higher Education %</td>
<td>&gt;50%</td>
<td>X1</td>
<td>17%</td>
</tr>
<tr>
<td>Age 50+ %</td>
<td>&gt;50%</td>
<td>38%</td>
<td>30%¹</td>
</tr>
</tbody>
</table>

² DigiD (from Digital ID) is a Dutch nationwide personal authentication system (sometimes called a digital passport) currently in use by different governmental institutions to verify citizens who use Web services of governmental organizations.
OBSTACLE 3: IMPACT ON DECISION MAKING

Is the citizen input actually used in the decision-making process? This question is quite difficult to answer for the investigated planning processes still are ongoing. In Maastricht, all the citizen reactions were analyzed by a person who had to distinguish “main trends,” leaving room for cherry-picking. These main trends were published in a concept-discussion cycling plan. This plan will be discussed with local stakeholder organizations, after which an implementation plan will be formulated. Maastricht plans to mirror this implementation plan once more to the original citizen input. In Zoetermeer, citizens could react on proposed plans in a virtual meeting, but it is unclear to what degree their comments affected decision making. In Helmond, voting was the most important opportunity to influence decision making, for the forum and the chats served mostly to inform people. However, the city council decided not to build the design with the most votes, but a combination of the designs for they received nearly the same amount of votes. Maastricht and Helmond as well as Zoetermeer state they use the application to get an idea of what the average person thinks and not directly to guide spatial changes. This clearly marks the limited impact of the citizen input on the decision-making process.

OBSTACLE 4: FEEDBACK MECHANISM

When using any of the applications, citizens cannot find information about the way their input might affect decision making or what feedback they can expect. As stated previously, Maastricht plans to mirror the implementation plan once more to the original citizen input. The people who registered and left their mailing addresses will be contacted to participate in the formal participation procedure of the cycling plan later in the planning process. In Helmond, feedback was guaranteed only when citizens posed questions on the forum. In Zoetermeer, the citizens did not receive feedback on their comments. Nevertheless, both developers and municipalities underline the risk of backfire if citizens do not feel their suggestions are taken seriously.

Discussion and Conclusion

E-participation has the potential to involve more citizens than does traditional participation meetings and seems to attract a different public. A user-friendly application that offers typical normative and instrumental participation functions can be seen as a precondition for an effective e-participation project. But, eventually, the participatory value of a project depends on the political will to utilize these functionalities and use the citizen input in decisions. Not all the investigated applications show the same suitability for participation. Both the E-spraak and Virtual Helmond applications provide little technical barriers and prove promising new channels for different functions of participation. Second Life proved a less suitable participation platform. Only recently, after the interviews have been conducted, Zoetermeer decided to stop Second Life Zoetermeer for the application was too difficult for many citizens and required a long installation and registration procedure (Van Rossum 2009). In all three cases, the translation from citizen input to decision making largely remains a black-box operation and citizens often do not receive the necessary feedback on the comments they made. Despite the claimed advantages and the technical possibilities to deepen participation with e-participation, local governments still hesitate to empower citizens.

The identified obstacles provide planners who implement e-participation with an overview of issues that can prove useful when starting a process. However, the current work of developers to improve participation in the applications might prove regretful if policy makers are not yet ready to involve citizens in decision making. The planning community should take responsibility for this problem and bring successful cases of e-participation to the attention of policy makers. A simple step to improve participation is by creating transparency: Inform citizens on how their input is used in the process and require mailing addresses to keep citizens involved in the process.

A blind spot still exists concerning the role of citizens in e-participation. Only one study performed a small survey among citizens (Carver et al. 2001). There is an urgent need to assess the position of citizens in a PSS. What citizens participate, how do they experience e-participation, and what limitations do they feel? The potential disadvantages of e-participation concerning authentication should receive attention. As face-to-face contacts are not possible, quasi-participation remains a risk until now. Experiments with digital authentication seem useful.

It is worth highlighting that this study focused on the rare municipalities that experiment with e-participation; additional research is necessary to investigate the considerations of the majority of the municipalities currently not applying e-participation.

About the Authors

Arjen Koekoek obtained a MSc in Land Use Planning (2006) at Utrecht University and a MSc in Geographical Information Science in 2008 at Wageningen University, both in The Netherlands. He is currently working as GIS-consultant at Geodan Next, where his interests include the use of GIS for interactive planning purposes.

Ron van Lammeren, originally educated as a landscape planner, obtained a PhD in Environmental and Agricultural sciences (1994) at Wageningen University, the Netherlands. His dissertation discusses the GIS support of the early 90’s Dutch regional planning. He is currently working as an associated professor in Geo Information Science at Wageningen University. His research focus is on geo-visualisation, PPGIS, and Interactive Location Based Services impacts on education and planning.
Dr. Guido Vonk is an independent researcher and consultant. The focus of his research work is on Urban Sustainability and Planning Support Systems. In his research, he investigates these systems from the viewpoint of their adoption by the planning community and their application in planning practice.

Acknowledgments

The authors wish to thank the five interviewed professionals, Mr. Van den Berg (municipality Helmond), Mr. De Vries (municipality Zoetermeer), Mr. Wahl’s (municipality Maastricht), Mr. van Eijk (Goudappel Coffeng), and Mr. Veth (Cebra) for their cooperation.

References


Footnotes
1 Van der Eijk and Bos (2007) estimate that the average user had a lower education based on the amount of spelling errors in the citizen input.
2 55+-years old instead of 50+-years old.
Using GIS to Contrast Perceived Versus Preferred Priorities for Brownfield Redevelopment in Worcester, Massachusetts

Claire W. Brill

Abstract: This paper compares the perceived priorities of decision makers with the stated desires of stakeholders concerning brownfield redevelopment in the city of Worcester, Massachusetts. Redeveloping brownfields—remnants from Worcester’s industrial past—is held as a critical strategy for the future of this city in central Massachusetts. However, the goals of this strategy vary across stakeholder groups. Key informants were surveyed regarding their perceptions of brownfield redevelopment. An inventory of Worcester brownfields was created from the Massachusetts Department of Environmental Protection records. A multicriteria evaluation was carried out using geographic information systems. This study found that while economic development and job creation are the apparent focus for one set of influential decision makers, protection of natural resources and public health are important to another group of stakeholders. The outcomes for each end-use objective were examined to determine whether the goals for each could be met while focusing on only one redevelopment strategy. Results reveal the land parcels that satisfy the objectives for only one group, both groups, or neither group. Previously redeveloped sites were evaluated based on the same criteria to determine the extent to which existing and preferred priorities had been satisfied with these revitalization efforts. The results show that focusing on perceived priorities will not accomplish preferred objectives for brownfield redevelopment in Worcester.

INTRODUCTION

Research Rationale

In recent years, brownfield redevelopment has become a strategy for sustainable land use and urban revitalization (BenDor and Metcalf 2005). The costs associated with not mitigating and redeveloping brownfields, apart from long-term health risks, include loss of tax revenues and a decrease in density of economic activity in urbanized areas (Meyer 2003). Urban sprawl into outlying green spaces, a hollow urban core, and redundant infrastructure are by-products of ignoring brownfields. Redeveloping brownfields remaining from Worcester’s industrial past is held as a critical strategy for the future of this city in central Massachusetts. With a finite supply of land within city boundaries, reclaiming brownfields is an important strategy for economic growth. Expansion of the tax base is necessary for the city to maintain the resources to meet the challenges inherent in an urban environment. Reclaiming brownfield sites will have a major impact on future economic growth and reclamation of neighborhoods (Kotsopoulos 2001). Worcester has had success in the past at redeveloping contaminated sites, such as the downtown hospital complex, often relying on public/private partnerships to advance projects. Brownfield redevelopment has been a priority issue within city government. The mayor has commissioned a Brownfields Roundtable to bring together local environmental, financial, and development professionals to continue the success (Nemeth 2005). The roundtable has lobbied successfully on behalf of brownfield legislation. However, the goals of the brownfield redevelopment strategy vary across stakeholder groups in the city. McCarthy (2002) noted that residents in neighborhoods with brownfields preferred recreational and community facilities and affordable housing as end uses for the properties. There is a division between community members and the government, whose primary goals for brownfield redevelopment include increased jobs and tax base. This research examines the difference between stakeholder priorities and addresses the question of whether perceived priorities serve preferred objectives for brownfield redevelopment in the city of Worcester.

Literature Review

Issues for consideration in brownfield redevelopment are documented in the literature. Categories of costs associated with brownfields, both related to cleaning them up and leaving them alone, are explored in Meyer (2003). Alberini et al. (2005) used pairwise comparisons of hypothetical brownfields with different policy mixes to explore the differences between those developers with experience in brownfields and those with none to determine which policies are attractive as incentives to promote redevelopment. The importance of an accurate inventory to help track the impact of redevelopment projects in the community was the impetus behind Coffin (2003) discussing the development of a database of brownfield properties. Moo-Young and Alattar (2003) included a site-selection survey used in Pennsylvania to rank brownfields for potential redevelopment. The criteria used relate to transportation access, infrastructure, zoning, size of parcel, and previous reuse efforts. The conclusion reached by Leigh (2002), that brownfield redevelopment rates in poverty neighborhoods are lower than those for nonpoverty neighborhoods, suggested the need to increase redevelopment in poverty neighborhoods to reduce a trend toward widening inequalities.
Researchers have utilized geographic information systems (GIS) for brownfield decision making. Boot (2001) emphasized data collection and a participatory process in establishing a GIS for brownfields, concluding that the development process should be iterative. A Brownfield Site Ranking Model, utilizing Smart Places’ ArcView extension, was created by Thomas (2002, Landscape and Urban Planning) for selecting sites for potential redevelopment. That research discussed the importance of knowing user needs in relation to the brownfields, obtaining site-specific data to differentiate between brownfields, and employing specific indicators and measurable criteria for ranking the sites. Rocco et al. (2002) provide a methodology for decision making regarding brownfield property reuse, including the kinds of data needed to make useful decisions. Risk profiles and factor weights are considered in the process through using GIS. Internet-based GIS applications were part of the strategy in redeveloping brownfields in Emeryville, California (Dayrit et al. 2002). Choosing target parcels for redevelopment in an inner city, based on particular objectives and criteria, was the focus of Simons and Salling (1995). The paper presented a methodology for using GIS as a decision support tool in the planning process. Milionis et al. (2000) applied GIS to the issue of brownfield redevelopment as they performed a risk assessment for human health and analyzed site suitability for a particular reuse using weighted scores.

The literature suggests that more than economic benefits are available with brownfield redevelopment. Broadening the revitalization focus to include a greater mix of land uses can help bring a higher quality of life to urban areas with brownfields. Creating green spaces out of brownfields can complement other types of redevelopment, such as the more typical commercial or industrial, on nearby properties (De Sousa 2006). Lack of integration between public health and the physical environment has been suggested as one reason behind the obesity epidemic in the United States. Redeveloping a brownfield in a neighborhood can improve the neighborhood and people’s health whether jobs or green spaces are created (Black 2000). Redeveloping brownfields into new housing is one strategy for reducing crime and incivilities in neighborhoods (Brown et al. 2004). Environmental justice and health disparities can be addressed through brownfield redevelopment depending on which sites are prioritized for redevelopment. As such, brownfield redevelopment is critical to urban revitalization. Public health benefits and successful neighborhood redevelopment can be achieved when public health is part of the decision-making process for brownfield redevelopment (Litt et al. 2002). Brownfield redevelopment is an effective smart growth policy. Both the environment of the urban core and outer greenfields can benefit from redeveloping brownfields (Greenberg et al. 2001).

A diverse redevelopment strategy must be pursued to allow for a mix of end uses. Most redevelopment projects focus on a single brownfield property; considering a wider area in the redevelopment strategy takes into account the benefits of off-site development and investment. Meyer (1998) holds that off-site benefits may add support to make projects more feasible. In addition, an area-wide perspective helps to connect brownfield policy to broader economic development efforts. McCarthy (2002) states that government initiatives should focus on redeveloping the least viable sites instead of spending money on sites more likely to be taken up by private developers. The most common brownfield sites in the United States, according to Johnson (2002), are the small abandoned lots that bring blight to the neighborhood. There is little economic incentive for cleaning up these sites, but big effects to the impoverished neighborhoods if these sites are cleaned up. Promoting the redevelopment of smaller brownfield sites that are scattered around the urban landscape should be the focus of local governments (Meyer and Lyons 2000). Improving infrastructure and services in affected neighborhoods may assist in drawing developers who would combine the smaller brownfield sites with surrounding properties for redevelopment. Thus, entire neighborhoods might be revitalized.

Measures by which redevelopment efforts are evaluated can impact project choices. Economic return has primarily been the measure of success in brownfield revitalization. De Sousa (2005) raised concerns that such a narrow emphasis ignores community goals. When outcome measurements primarily detail economic factors such as tax revenue increases or job creation, social and environmental outcomes are viewed as less important. The National Brownfields Environmental/Community Caucus (1999) recommends broadening the definition of success to include improvement in public health, creation of ownership opportunities for the community, and enhancement of the quality of life in neighborhoods containing brownfields. McCarthy (2002) recognizes the difficulty of connecting redevelopment to broader goals because of the non-economic factors that governments cannot completely control. It is hoped that a common desire for brownfield redevelopment would turn these challenges into opportunities.

Research Location
The study area under consideration is the city of Worcester, Massachusetts. Located in central Massachusetts, Worcester has a population of approximately 180,000, making it the second largest city in New England. It covers about 98 square kilometers. Worcester contains several institutions of higher learning and a burgeoning biotechnology industry. The history of the city includes many innovations in manufacturing. It is the legacy of this industrial past that is evident in the many brownfields in Worcester. City government considers brownfield redevelopment a key issue. The mayor’s Brownfields Roundtable exists to promote redevelopment and influence legislation. This paper examines the current priorities for brownfield redevelopment and whether the variety of views of all stakeholders is represented in those priorities.

METHODS
Data
Several sources supplied the data for this study. The city of Worcester GIS Section of the Technical Services Division provided the majority of the data related to city infrastructure. The Office of Geographic and Environmental Information (MassGIS), Com-
monwealth of Massachusetts Executive Office of Environmental Affairs provided data related to environmental justice, aquifer location, and colleges and universities (http://www.mass.gov/mgis/massgis.htm). Brownfield information was obtained through the Massachusetts Department of Environmental Protection (MassDEP) Web site (http://www.mass.gov/dep/cleanup/sites/sitelook.htm). Additional data associated with crime and current development was derived from facts on the city of Worcester official Web site (http://www.ci.worcester.ma.us). The Census Bureau was the source for some demographic statistics (http://factfinder.census.gov/home/saff/main.html?_lang=en). Data were initially processed in the ArcGIS software. All layers were projected in NAD 1983 State Plane Massachusetts Mainland with a distance unit of meters.

For this study, brownfields are defined as properties with a Release Tracking Number (RTN) as assigned by MassDEP. The 485 brownfields being considered in this study are pictured in Figure 1.

**PROCEDURE**

**Survey**

The first stage of the research was to administer a survey. The anonymous survey (see Figure 2) was given to the volunteer participants of the mayor’s Brownfields Roundtable at a regular monthly meeting.

Dedicated government officials, legal professionals, and environmental and regulatory experts involved with the brownfield issue in the city of Worcester are part of the roundtable. The survey was conducted at a meeting with average attendance. Approximately 50 percent of those in attendance completed the survey, which resulted in eight responses. The survey sample, albeit small, was representative of those most knowledgeable about brownfield redevelopment in the city. The survey gauged perceptions regarding brownfield redevelopment priorities in Worcester. Respondents were asked to rate 11 factors in two different ways. First, factors were rated, on a scale of 1 to 10, according to their perceived current influence on redevelopment in Worcester. Secondly, the same factors were rated according to what influence the respondents preferred. Additional stakeholders were interviewed to understand their priorities for brownfield redevelopment in Worcester. Academic researchers, an environmental activist, and government official not involved in the mayor’s Brownfields Roundtable shared their views on redevelopment priorities.

**GIS Layers**

Next, survey factors were translated into GIS layers. The literature and expert opinion suggested factors to include in the study.

---

**Figure 1.** Brownfield parcels (n = 485) in Worcester, Massachusetts, prioritized in this study

**Figure 2.** Survey instrument administered to mayor’s Brownfields Roundtable
The literature documents various methods for organizing data requirements. The siting guidelines suggested by Thomas (2002, *Landscape and Urban Planning*) reflect social and cultural criteria, including neighborhood cohesiveness and community education and involvement. For this study, the community development corporations and colleges layers represent this organization and education criteria. Milonis et al. (2000) collected layers in thematic categories, including parcel base map, zoning, demographics, and contamination and groundwater data. Categories for data in reuse site characterization undertaken by Rocco et al. (2002) pertained to land-use and economic issues as well as community and neighborhood needs and environmental risk assessment. Transportation infrastructure and compatibility to master plans were part of the criteria in the research completed by Thomas (2002, *Environmental Practice*). Interstate access, railroad locations, and cultural and development hotspots were the corresponding criteria for this research. Layers were chosen to represent the range of criteria from economic factors to quality-of-life issues. All GIS layers were created with publicly accessible data. Thus, some GIS layers, such as aquifer location, were recommended by experts as proxies for factors because of the readily available nature of the data. A list of layers can be viewed in Table 1.

Table 1. GIS layers that represent factors in the survey

<table>
<thead>
<tr>
<th>Factor in Survey</th>
<th>GIS Layer(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest by an investor</td>
<td>Parcel size, location of colleges and universities</td>
</tr>
<tr>
<td>Economic development</td>
<td>Vacant commercial/industrial and residential parcels, abandoned buildings, tax title properties</td>
</tr>
<tr>
<td>Proximity to major roads or highways</td>
<td>Interstate exits, railroads</td>
</tr>
<tr>
<td>Severity of contamination</td>
<td>MassDEP status codes</td>
</tr>
<tr>
<td>Current development in surrounding area</td>
<td>Cultural and development hotspots</td>
</tr>
<tr>
<td>Community support</td>
<td>Community development corporations, aquifer location</td>
</tr>
<tr>
<td>Public health</td>
<td>Residential, commercial/industrial zoning</td>
</tr>
<tr>
<td>Zoning</td>
<td>Water features, parks, community gardens</td>
</tr>
<tr>
<td>Protection of natural resources</td>
<td>Environmental justice zones, schools, children and elderly</td>
</tr>
<tr>
<td>Adjacency to vulnerable populations</td>
<td></td>
</tr>
<tr>
<td>Crime prevention</td>
<td>Percent of total incidents per police statistical area</td>
</tr>
</tbody>
</table>

Initial data processing occurred in ArcGIS software. Prepared vector layers were imported into Idrisi' Andes. Vector files were rasterized. A resolution of 20 m² was chosen based on sizes of known brownfields; a majority of brownfields would be larger than one pixel with this resolution. Additional processing of the raster layers, which included reclassing values or calculating distance from features, prepared the data for multicriteria evaluation.

**MULTICRITERIA EVALUATION**

Multicriteria evaluations were executed in Idrisi. First, raster layers were standardized on a 0 to 255 scale, with 255 indicating highest suitability. Standardization methods included using fuzzy set membership functions (Jiang 2000). Control points for these functions were derived from various sources. A spatial join with previously redeveloped brownfields was used to determine distances to features, such as highway exits, to use for control points. Relevant literature suggested important cutoffs for parcel size and contamination severity, for example. Massachusetts water-supply regulations informed the choice of control points for the public-health criterion.

Factors were weighted using a pairwise comparison process (Eastman 2006). A factor was scored based on its relative importance to each other factor. Survey ratings guided the decisions made in the pairwise comparison process. Weights were calculated for the two objectives: perceived redevelopment practices and preferred redevelopment practices. Weights derived using perceived influence ratings are listed in Table 2. The weights derived using the preferred influence ratings are documented in Table 3.

Table 2. Weights assigned to each factor based on perceived influence survey results

<table>
<thead>
<tr>
<th>Survey Factor</th>
<th>Survey Ranking</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest by investor</td>
<td>1</td>
<td>0.1810</td>
</tr>
<tr>
<td>Economic development, perceived practice</td>
<td>2</td>
<td>0.1608</td>
</tr>
<tr>
<td>Proximity to transportation</td>
<td>3</td>
<td>0.1365</td>
</tr>
<tr>
<td>Severity</td>
<td>4</td>
<td>0.1229</td>
</tr>
<tr>
<td>Current surrounding development</td>
<td>5</td>
<td>0.0961</td>
</tr>
<tr>
<td>Community support</td>
<td>6</td>
<td>0.0890</td>
</tr>
<tr>
<td>Public health</td>
<td>7</td>
<td>0.0768</td>
</tr>
<tr>
<td>Commercial/industrial zoning</td>
<td>8</td>
<td>0.0603</td>
</tr>
<tr>
<td>Protection of natural resources</td>
<td>9</td>
<td>0.0409</td>
</tr>
<tr>
<td>Adjacency to vulnerable populations</td>
<td>10</td>
<td>0.0232</td>
</tr>
<tr>
<td>Crime</td>
<td>11</td>
<td>0.0126</td>
</tr>
</tbody>
</table>
Table 3. Weights assigned to each factor based on preferred influence survey results

<table>
<thead>
<tr>
<th>Survey Factor</th>
<th>Survey Ranking</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic development, preferred practice</td>
<td>1</td>
<td>0.1482</td>
</tr>
<tr>
<td>Interest by investor</td>
<td>2</td>
<td>0.1399</td>
</tr>
<tr>
<td>Current surrounding development</td>
<td>3</td>
<td>0.1192</td>
</tr>
<tr>
<td>Adjacency to vulnerable populations</td>
<td>4</td>
<td>0.1062</td>
</tr>
<tr>
<td>Community support</td>
<td>5</td>
<td>0.0974</td>
</tr>
<tr>
<td>Proximity to transportation</td>
<td>6</td>
<td>0.0974</td>
</tr>
<tr>
<td>Severity</td>
<td>7</td>
<td>0.0869</td>
</tr>
<tr>
<td>Protection of natural resources</td>
<td>8</td>
<td>0.0678</td>
</tr>
<tr>
<td>Public health</td>
<td>9</td>
<td>0.0678</td>
</tr>
<tr>
<td>Residential zoning</td>
<td>10</td>
<td>0.0397</td>
</tr>
<tr>
<td>Crime</td>
<td>11</td>
<td>0.0296</td>
</tr>
</tbody>
</table>

These weights represent the importance assigned to each factor based on survey ratings. Standardized factors were aggregated with the derived weights through a weighted linear combination for each objective. The average of all pixel suitability scores within each brownfield was extracted and assigned as the suitability score for the entire brownfield. Suitability scores for brownfield redevelopment were calculated twice, first according to the perceived practice objective (see Figure 3) and second based on the preferred practice objective (shown in Figure 4).

Twelve previously redeveloped brownfield parcels also were evaluated. Each was assigned a suitability score based on the multicriteria evaluations for perceived and preferred objectives to use in analysis.

**DISPLAY AND ANALYSIS**

Display of results was accomplished in ArcGIS. Two maps of suitability scores for each brownfield were created. One map contained the scores for the perceived practice objective and the other for the preferred practice objective. Brownfields were ranked based on suitability scores, with 485 indicating the highest suitability for redevelopment and 1 indicating the lowest suitability score.

Various analyses of the results were completed. Ranks from the perceived practice were plotted against ranks from the preferred practice to test for associations between the ranks. The suitability scores for previously redeveloped brownfields were compared to the scores of existing brownfields to understand how the different objectives were met by the revitalized properties. A regression analysis was performed to determine which factors had the most influence on the difference in ranks between the objectives. The difference between ranks for each brownfield (preferred ranking minus perceived ranking) was the dependent variable. The predictor, or independent, variables in the regression were the standardized factor maps.

Figure 3. Multicriteria evaluation results for perceived practice objective. Black indicates areas that are not brownfields. Larger values indicate higher suitability for redevelopment.

Figure 4. Multicriteria evaluation results for preferred practice objective. Black indicates areas that are not brownfields. Larger values indicate higher suitability for redevelopment.

Low Suitability

High Suitability

Low Suitability

High Suitability

Low Suitability

High Suitability

Low Suitability

High Suitability

Low Suitability

High Suitability
The survey results show that perceived current practice and preferred practice for brownfield redevelopment are different. A bar chart (see Figure 5) summarizes the survey results, illustrating the average rating each factor received for perceived and preferred influence on redevelopment.

Factors related to economic development received the highest ratings according to their perceived influence on brownfield redevelopment. Factors related to quality of life, such as public health and protection of natural resources, received a lower average rating based on perceived current brownfield redevelopment practice. Quality-of-life factors received a higher average value when respondents rated based on preferred practice for redevelopment. Economic factors still were among the top rated influencers for this objective (see Table 3). However, the difference in ratings between all factors was less for preferred than for perceived practice.

Conversations with other stakeholders revealed an even stronger dichotomy. While economic development and job creation are the apparent exclusive focus for one set of influential decision makers, protection of natural resources and public health are of primary importance to another group of stakeholders. A plot of ranks revealed a lack of tight correspondence of the ranks of brownfields based on perceived practice and the ranks of brownfields based on preferred practice (see Figure 6). The top 20 ranked brownfields for each objective showed two brownfields overlapped the objectives.

Maps of the top 20 brownfields were created (shown in Figures 7 and 8).

The top 20 brownfields according to perceived practice generally were larger parcels along interstate corridors (see Figure 8). The top 20 brownfields according to preferred practice were clustered together farther from the interstate corridors (see Figure 9).

A statistical analysis illuminated the top factors influencing the difference in rankings between objectives. The difference in ranks (preferred minus perceived) was regressed against all factors. The first four factors added into the equation in a forward stepwise
regression ($\alpha = 0.05$) were: (1) transportation factor for preferred practice ($\beta = 0.671$), (2) vulnerable populations ($\beta = 0.326$), (3) residential zoning ($\beta = 0.300$), and (4) economic development factor for perceived practice ($\beta = -0.234$). A positive $\beta$ coefficient means that an increase in the factor score causes an increase in the difference between ranks. A negative $\beta$ coefficient means that an increase in the factor score causes a decrease in the value of preferred minus perceived rank.

Suitability scores for previously redeveloped brownfields were derived using the same multicriteria evaluations. The 12 previously redeveloped brownfield parcels were overlaid with the top ranked brownfields prioritized in this study (see Figure 10).

Ranks received by the previously redeveloped brownfields based on perceived practice were plotted against ranks based on preferred practice (shown in Figure 6). The redeveloped brownfield with the highest suitability score based on the perceived objective ranked in the top 20 (472 of 485). None of the previously redeveloped brownfield sites received a suitability score based on preferred practice that was high enough to rank in the top 20; the previously redeveloped brownfield with the highest suitability score based on the preferred objective ranked 280 of 485.
DISCUSSION

Interpretation

Brownfield redevelopment has varying goals depending on the stakeholder. The U.S. Conference of Mayors (2003) report, which lists the most important benefits of brownfield redevelopment as increased tax base, job creation, neighborhood revitalization, and environmental protection, reflects this diversity. The perceived practice in the city of Worcester leans more to economic development and job creation. Stakeholders in the city do desire that quality-of-life issues be addressed through brownfield redevelopment. Yet, the current focus will not achieve both economic development and quality-of-life objectives. The results show that highly ranked sites based on preferred objectives are not associated with highly ranked sites based on perceived objectives. Considering Figure 5, however, average ratings of factors were not in complete opposition for the different objectives. It appears that increasing the influence of noneconomic issues, without diminishing the role of economic factors, will help to achieve preferred objectives.

Further Research

Multicriteria evaluation in GIS can be used for focusing stakeholders on the issue. As with all multicriteria evaluations, this is the first step in the decision-making process. With the results of this research, the next step would be to get stakeholders together for discussion. Many questions should be raised. How should brownfields be defined? Which factors should be represented with different data? Are there factors that are missing? Are the factors standardized and weighted appropriately? What kind of trade-off exists between different factors? Answers to these questions will guide the next steps for this research.

CONCLUSIONS

This research has shown through using GIS in a multicriteria evaluation that different results are achieved with differing brownfield redevelopment perspectives. Exclusive focus on perceived priorities would not meet the broader list of preferred objectives for brownfield redevelopment in the city of Worcester. However, there is some common ground. Economic factors were highly ranked in both objectives and two properties received a top 20 ranking from each perspective. While economic factors were important to both perceived and preferred objectives, quality-of-life factors held increased weight in the preferred objective. In an ideal world, every brownfield would be brought back into productive use. Incorporating diverse factors when deciding in which brownfields to focus finite resources may help promote benefits enjoyed across the community. Brownfield redevelopment is a critical issue in postindustrial Worcester. However, the goals of the brownfield redevelopment strategy vary across stakeholder groups in the city. If the government will represent all its constituents, it must consider a broader set of criteria for prioritizing properties. Perhaps the mayor’s Brownfields Roundtable could take the lead, as it has on other issues in the past, to focus attention on ways to satisfy the preference to improve the quality of life while the existing economic concerns also are met. Perhaps new partnerships between the mayor’s Brownfields Roundtable and other key players in the city can help to achieve brownfield redevelopment results that would be considered successful from the differing objectives.

Acknowledgments

This research would not have been possible without the generous sharing of data by the former city of Worcester GIS Manager, Shane D. White. The data resources of the Office of Geographic and Environmental Information (MassGIS), Commonwealth of Massachusetts Executive Office of Environmental Affairs enhanced this research. Clark Labs facilitated this work by creating the GIS software Idrisi. ESRI is the creator of ArcGIS, another GIS software essential to this research. Input from the mayor’s Brownfields Roundtable of the city of Worcester and various other key informants in the private and public sector was greatly appreciated.

About the Author

Claire W. Brill received her MA in Geographic Information Sciences for Development and Environment from Clark University in May of 2008. She earned a BS in Atmospheric Science from Purdue University and an MA in Teaching from Fairfield University. Currently, the author is a GIS analyst for the city of Worcester, Massachusetts. She also is Chair of the GIS Subcommittee of the city of Worcester mayor’s Brownfields Roundtable.

Corresponding Address:
94 Acushnet Avenue
Worcester, MA 01606
Phone: (508) 853-4638
cwbrill@alumni.clarku.edu

References


Mark Your Calendar!
Mark Your Calendar!

URISA’s 48th Annual Conference & Exposition

September 28 – October 1, 2010

Orlando, Florida