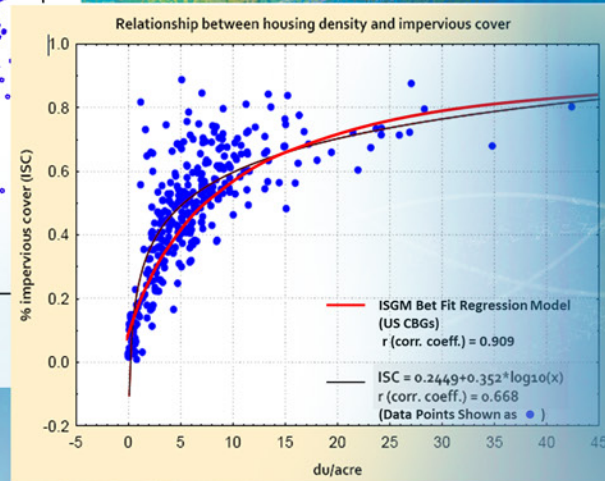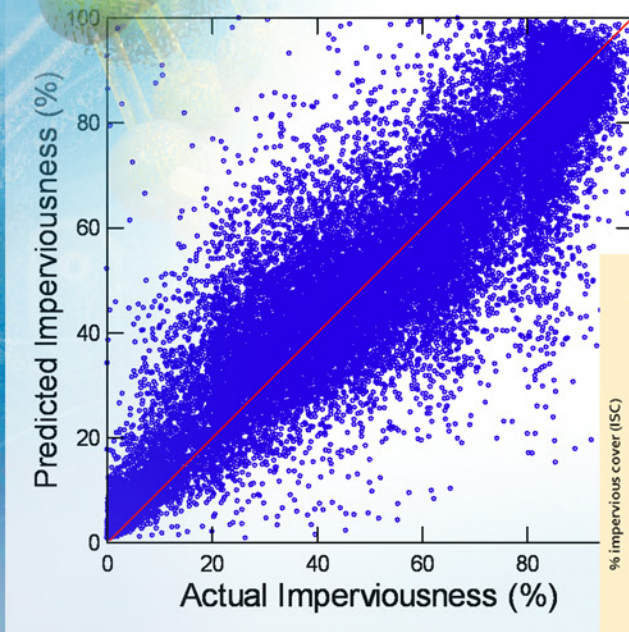# Journal
of the Urban and Regional Information Systems Association

URISA

Smart Eco-path Finder for Mobile GIS Users

A Qualitative Framework for Evaluating Participation on the Geoweb

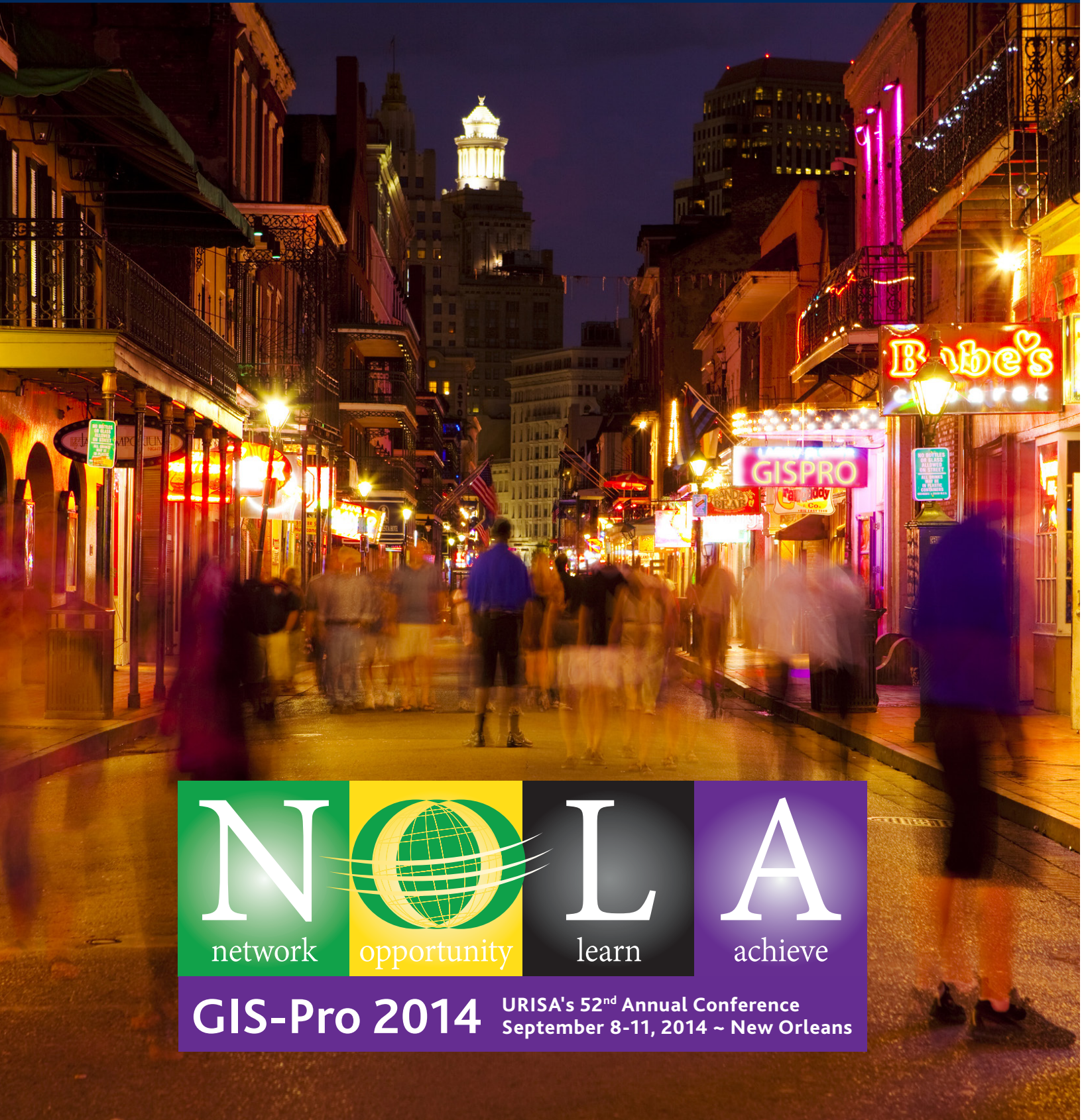A Place-based Tool for Assessing Cumulative Impervious Surface Outcomes of Proposed Development Scenarios

The Role of Collaboration in Spatial Data Infrastructures

Papers of Practice, Technical Reports and Industry Notes

Building the School Attendance Boundary Information System (SABINS): Collecting, Processing, and Modeling K to 12 Educational Geography

Improving Cadastre: Development of a Workflow Prototype Utilizing ESRI's Parcel Fabric

MARK YOUR CALENDAR!

NOLA

network   opportunity   learn   achieve

GIS-Pro 2014

URISA's 52nd Annual Conference
September 8-11, 2014 ~ New Orleans

URISA.ORG

CONTENTS

URISA Journal

EDITORIAL BOARD

Smart Eco-path Finder for Mobile GIS Users

Ko Ko Lwin and Yuji Murayama

Abstract: *The increasing popularity of the Internet and user-friendly Web-based GIS applications such as Google Maps/Earth and Microsoft Bing Maps Platform have made GIS an integral part of life today for finding the nearest facilities, driving routes, and so on. However, choosing an ecofriendly walking route is a big challenge for local residents because of the lack of GIS analytical functions and environmental data available online. Although analysis of route paths has been widely used in GIS applications, the integration of green factors with the analysis of the route path is still lacking in the GIS arena. In this paper, we have presented an integrated methodology, remote sensing, GIS and spatial Web technology, for identifying an ecofriendly walking route (Smart Eco-path Finder) by providing Web-based GIS analytical functions using Tsukuba City in Japan as a case study. This Web-based, ecofriendly route pathfinder enables users to choose a walking route for green exercise by using their smartphones or ultramobile PCs (UMPCs).*

INTRODUCTION

Walkability captures the proximity between functionally complementary land uses (live, work, and play) and the directness of a route or the connectivity between destinations (Forsyth and Southworth 2008, Moudon et al. 2006). A walk score is an indicator of how "friendly" an area is for walking. This score is related to the benefits to society in terms of energy savings and improvement in health that a particular environment offers its residents. For example, a recently developed walk score Web site uses Google Maps, specifically Google's Local Search API (application programming interface), to find the stores, restaurants, bars, parks, and other amenities within walking distance of any address that is entered. Walk score currently includes addresses in the United States, Canada, and the United Kingdom. The algorithm behind this score indicates the walkability of a given route based on the fixed distance from one's home to nearby amenities. The number of amenities found nearby is the leading predictor of whether people will walk rather than take another travel mode. However, evaluating walkability is challenging because it requires the consideration of many subjective factors (Reid 2008). Moreover, all technical disciplines related to walkability have their own terminology and jargon (Abley 2005).

During the urban and regional planning processes, the green spaces and the environmental quality of neighborhoods are important factors that affect human health. Fortunately, green spaces and neighborhood environmental quality can be improved through proper urban management. Thus, epidemiological studies have explored the relationship between access to nature and health. For example, a study in Sweden by Grahn and Stigsdotter (2003) demonstrated that the more often one visits green areas, the less often one reports stress-related illness. One epidemiological study performed in the Netherlands (Maas et al. 2006) showed that residents of neighborhoods with abundant green spaces tended, on average, to enjoy better general health. Another possible mechanism relating nature to health occurs

during social interactions and social cohesion. Several studies conducted in Chicago suggest that green spaces, especially trees, may facilitate positive social interactions between neighboring residents (Kweon, Sullivan, and Wiley 1998). Moreover, Pretty et al. (2007) summarized the effects of ten green exercise case studies (including walking, cycling, horseback riding, fishing, canal boating, and conservation activities) in four regions of the United Kingdom on 260 participants. They determined that green exercise (i.e., exercise in a green area) led to significant improvements in self-esteem and in total mood. The results were not affected by the type, intensity, or duration of the green exercise. Therefore, in many parts of the world, current urban planning activities are shifting toward a focus on "green" living. Many cities around the world now are developing integrated solutions to major environmental challenges and are transforming themselves into more sustainable and self-sufficient communities (Dizdaroglu, Yigitcanlar, and Dawes 2009).

On the other hand, GIScience provides theory and methods that have the potential to facilitate the development of spatial analytical functions and various GIS data models, which improve the building of sophisticated GIS systems. Among them, the GIS road network data model is important for solving the problems in urban areas, such as transportation planning, retail market analysis, accessibility measurements, service allocation, etc. There are several network models in GIS, such as river networks, utility networks, and transportation networks or road networks. Understanding the road network patterns in urban areas is important for human mobility studies, because people live and move along the road networks. A network data model allows us to solve daily solutions, such as finding the shortest or quickest path between two locations, looking for the closest facilities within a specific distance, and estimating driving time. Although many network models are conceptually simple, they are mathematically complex and require computational resources to model the problem (Lwin and Murayama 2011).

Perhaps the most exciting area of computer system development continues to be in handheld devices such as netbooks, tablet PCs, and smartphones. A smartphone is a mobile phone that offers more advanced computing ability and connectivity than does a phone with contemporary features. They are recognizably efficient in form factor (size, shape, weight, etc.), chip type, internal storage capacity, battery life, and operating system compared to desktop computers. Along with these hardware and communication technology developments, the emergence of user-friendly Web-based GIS applications such as Google Maps/Earth and Microsoft Bing Maps Platform have made GIS an integral part of life today for finding the nearest facilities, driving routes, and so on. Nowadays, GIS goes everywhere, desktop to laptop to netbook to smartphone and professionals to nonprofessionals (i.e., expert users to public)—we are now in a "geospatially enabled society." At least our mobile phones tell us where we are and what kinds of facilities are available near us.

The quality of ecofriendly living places or walking routes can be measured by an indicator of walkability index or score. Although most walk score calculations are based on distances between home and public facilities, an ecofriendly walk score calculation is based on green spaces (i.e., location of home or a walking route with green spaces). The higher the score, the better the environmental quality (i.e., ecofriendly) for living or doing green exercise. Finding the ecofriendly places reduces the cooling and heating demand, improves the air quality, reduces storm-water runoff, enriches urban biodiversity and urban agriculture, reduces urban heat island effect, contributes to a carbon-neutral architecture, aesthetically improves the cities' skylines, and creates a positive economic impact (Roehr and Laurenz 2008)—and walking through ecofriendly routes (i.e., ecowalkability) allows local residents to use the lowest-cost transportation mode (i.e., their feet) to reduce energy consumption and improve physical conditions.

Choosing an ecofriendly place to live or for a walking route, however, is a big challenge for local residents because of the lack of GIS analytical functions and environmental data available online. Although an analysis of route paths has been widely used in GIS applications, the integration of green factors with the analysis of the route path still is lacking in the GIS arena. Given the great

Table 1. Data, descriptions, and applications of their use

Data and Source	Description	Purpose
ALOS AVNIR-2 (Japan Aerospace Exploration Agency JAXA)	• Band 3 (Red: 0.61~0.69 μm) • Band 4 (Infrared: 0.76~0.89 μm) • ten-m spatial resolution at Nadir • Raster in GeoTIFF format	• To compute Normalized Different Vegetation Index NDVI • To delineate greenness spaces • To convert binary green image • To compute the greenness score
Building footprints (Zmap-TOWNII product from ZENRIN Company)	• Building footprints including building name, parcel number, and number of floors • Polygon in an ESRI shapefile	• To integrate with administrative boundary data and construct a database of residential addresses • To create masks on vegetated areas
Administrative boundary (Zmap-TOWNII product from ZENRIN Company)	• Administrative boundary including name • Polygon in an ESRI shapefile	• To integrate with building footprints and create a database of residential addresses • To calculate the greenness score by administration zone
Road centerlines (Geospatial Information Authority of Japan)	• Road centerlines with major road names • Line in an ESRI shapefile	• To build a road network model • To measure network distances between a user-defined point and locations of facilities • To compute a greenness score for each road segment • To perform an analysis of the shortest or greenest route
Facility locations (iTownpage from NTT, Nippon Telegraph & Telephone Corp.)	• Business name, address, category, subcategory, business contents, telephone number, URL, etc. • Comma separated value (CSV) format	• To convert a point layer for facilities • To find desirable and available facilities by a user-defined search distance
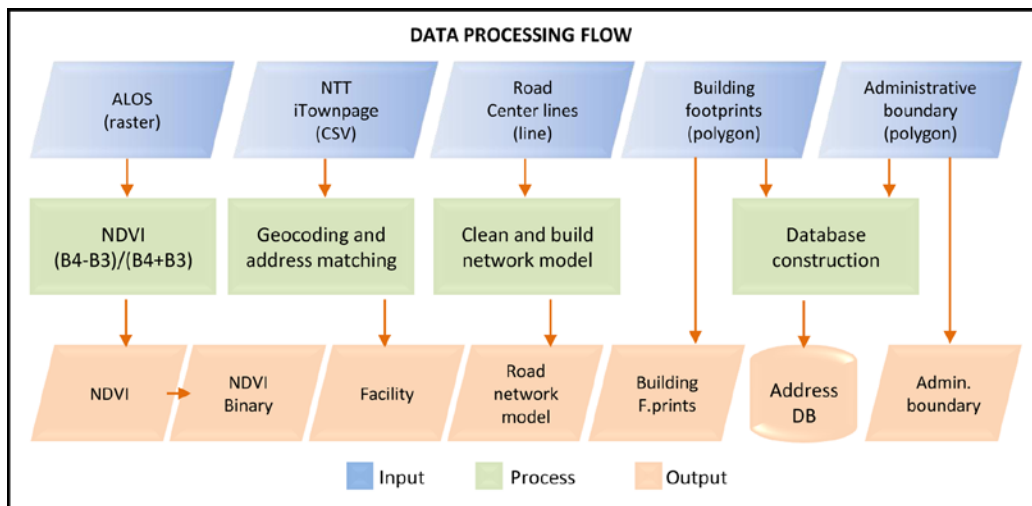
Figure 1. Data-processing workflow (Source: Lwin and Murayama 2011)

interest for walking activities combined with green spaces, the purpose of this paper is to develop an integrated methodology (remote sensing, GIS, and spatial Web technology) to identify the ecopath or greenest routes for mobile GIS users using their smartphones or UMPCs. The Smart Eco-path Finder project is further development of our previous project named "Ecofriendly Walk Score Calculator: Choosing a Place to Live with GIS" for local residents to make spatial decisions for choosing ecofriendly living places in Tsukuba City (Lwin and Murayama 2011), which has been tested and evaluated by local residents especially real estate agencies, nonprofit organizations (NPOs), local residents, and university students. A handful of university students were requested to use the greenest pathfinder on their smartphone or tablet PC to find the greenest route while they walk or exercise. Overall, the system was evaluated favorably by real estate agencies, researchers, and university students in Tsukuba City.

DATA PREPARATION FOR SMART ECO-PATHFINDER

The development of the Smart Eco-path Finder is based on spatial Web technology (i.e., WebGIS), which was built on both advantages of GIS and Internet technology, also referred to as GIScience and Technology (GIST). In this section, we will discuss the data, purposes, and processing steps for the establishment of the Smart Eco-path Finder system. Table 1 shows the data and sources, description of data, and purpose to use in this project.

Figure 1 shows the workflow for the data processing and output of data to be used in the Eco-friendly Walk Score Calculator.

The following data preprocessing steps were involved in this system.

CREATION OF A BINARY GREEN IMAGE FROM ALOS SATELLITE DATA

ALOS includes an optical sensor known as the Advanced Vis-

ible and Near Infrared Radiometer type 2 (AVNIR-2) with high spatial resolution (ten m at nadir) composed of four multispectral bands (i.e., three bands in the visible region and one band in the near-infrared region). The ALOS-derived normalized difference vegetation index (NDVI; NDVI = (NIR-RED)/(NIR+RED)) is computed using a visible red band (RED, Band 3: 0.61 - 0.69 μm) and a near-infrared band (NIR, Band 4: 0.76 - 0.89 μm) acquired from vegetation growing seasons. This NDVI index shows the degree of vegetation (intensity) represented as pixel values between 0 and 256, which are stretched from their original values between -1 and 1.

After stretching, this NDVI image was resampled to a 5 m spatial resolution using the cubic convolution method. Cubic convolution is a popular method for image interpolation, which is the process of defining a spatially continuous image from a set of discrete samples. Image interpolation is fundamental to many digital image-processing applications, particularly in operations requiring image resampling, such as scaling, registration, warping, and correction for geometric distortions. Interpolation commonly is implemented by convolving an image with a small kernel for the weighting function. Popular methods of interpolation by convolution include nearest neighbor interpolation, bilinear interpolation, cubic B-spline interpolation, and piecewise-cubic convolution (Lehmann et al. 1999).

To separate vegetated and nonvegetated spaces, we set the threshold at 113 of NDVI pixel values by comparing two images (i.e., one from a 67-cm RGB-321 True Color Ortho-image and 5 m resampled ALOS NDVI image) using View > Link/Unlink Viewers > Geographical function in ERDAS Imagine commercial remote-sensing software. After this step, the intensity image is converted into a binary green image (one for vegetated area and zero for nonvegetated area). The main purpose of this conversion is to identify the vegetated areas rather than the vegetation intensities, which vary from season to season. The binary green image also reduces the data size and the required computational time. This procedure is especially suitable for Web-based GIS

in which the network and computational resources are limited. Vegetated areas included trees, bush lands, grasslands, and paddy fields. Nonvegetated areas include buildings, parking lots, bare lands, rivers, and lakes.

ROAD NETWORK DATA MODEL AND CALCULATION OF GREENNESS SCORE

Road centerline data were acquired from the Geospatial Information Authority of Japan. This dataset, however, does not cover all the small streets in the city. Therefore, any small streets missing are digitized based on Zmap-TOWNII data. Moreover, additional sides of the major roads are required to digitize again, for major roads are more than 10 m in width (see Figure 2) and because we compute a greenness score for each road segment within a 10 m buffered area.
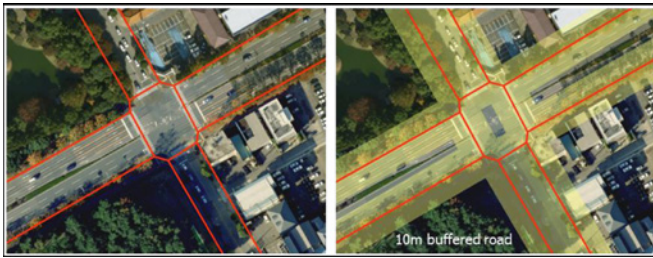
Figure 2. Additional side roads are required to digitize for the greenness score calculation

Following this, we add a 10 m buffer to both sides of the road and compute the greenness score based on the binary green image (see Figure 3) for each road segment using equation (1).
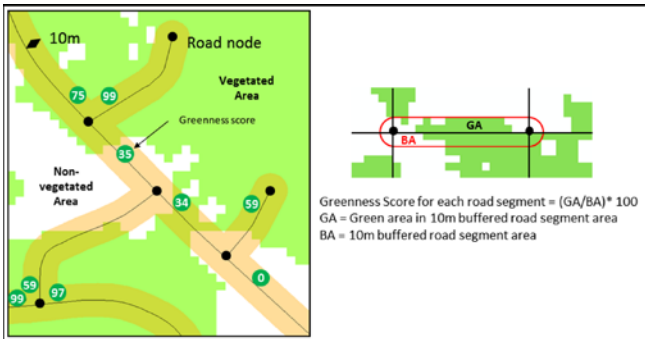
Figure 3. Calculation of greenness score for each road segment within the ten m buffered road

Calculation of greenness score for each road segment:
Greenness score for each road segment = (GA/BA)* 100
    (1)
GA = green area in 10 m buffered road segment area
BA = 10 m buffered road segment area

Next, we build a topological road network model using VDS Road Network Builder provided by VDS Technologies. Using this process, we set up the greenness score attribute field as a weighted factor to compute the shortest or greenest route between the points. The shortest route was computed based on road distance, and the greenest route was computed based on road distance and the greenness score whose value is between 0 and 100.

CONSTRUCTION OF RESIDENTIAL ADDRESS DATABASE

For this case study, a database of residential addresses is created from a combination of administrative boundary and building footprint datasets. Building footprint datasets are useful for estimating building populations (Lwin and Murayama 2009) because such data contain rich attributes including building number, number of floors, and building name. Unlike other countries, most addresses in Japan are based on a block-by-block system (prefecture block, city block, ward block, ownership block, etc.). The address does not contain a street or road name; instead, it is expressed by a sequence of blocks. For this study, we constructed the address database by performing an intersection function between these block layers. We separated the addresses into two parts: the main block and the subblock. The main block represented the smallest administrative unit, and the subblock represented the smallest land unit. For example, in the case of Kasuga 3-15-23, Kasuga 3 was constructed from an administrative boundary block, and 15-23 was constructed from the smallest land unit. The purpose of the address database is to locate the place in a user-friendly way and to avoid any problems with mistyping when performing an address search. Although this approach is not appropriate for large land blocks (e.g., factories, schools, and hotels), users still can locate their positions by clicking on a map and getting the X, Y coordinates.

Location of Public Facilities

We use iTownpage data, which were downloaded from the Nippon Telegraph & Telephone Corp. (NTT) Web site. These data include the business name, type, category, content, address, telephone number, and other information in a comma separated value (CSV) format. The iTownpage Web site supports the everyday lives and business activities of visitors and expatriates in Japan and people living overseas by enabling users to search for information about stores and businesses via the Internet. These CSV data were converted into ESRI point features using commercial geocoding software with an accuracy at the building level. These NTT iTownpage data can be used to separate the residential and nonresidential buildings and retail market analysis. Here we used it for building a public facility list in order to choose the place by facility name as the walking target.

IMPLEMENTATION

System Overview
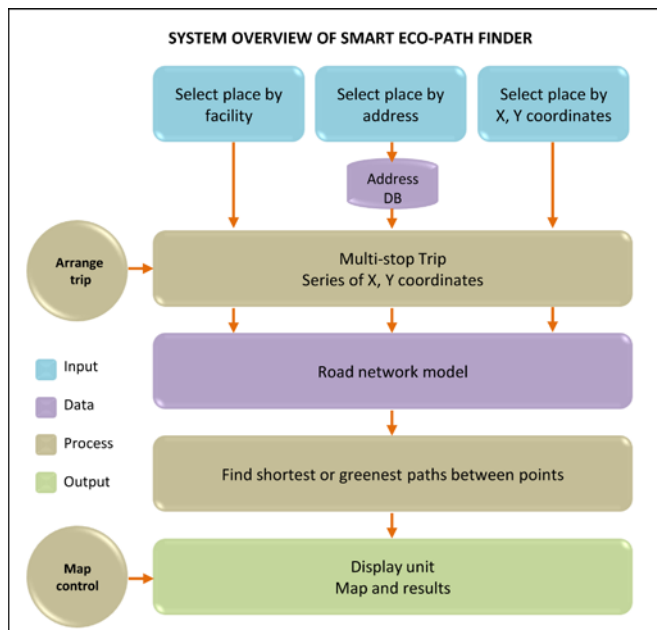
Figure 4. System overview of Smart Eco-path Finder

DEVELOPMENT PLATFORM

Figure 4 shows the overall system design; we have implemented this system called "Smart Eco-path Finder" based on Microsoft ASP.NET with an AJAX Extension and VDS Technologies (Web Mapping Components for ASP.NET). ASP.NET is a Web applica-tion framework marketed by Microsoft that programmers can use to build dynamic Web sites, Web applications, and XML Web services. AJAX (asynchronous JavaScript and XML) is a group of interrelated Web-development techniques used on the client side to create interactive Web applications. With AJAX, Web applica-tions can asynchronously retrieve data from the server in the back-ground without interfering with the display and behavior of the existing page. The use of AJAX techniques has led to an increase in interactive and dynamic interfaces on Web pages. AspMap for .NET from VDS Technologies is a set of high-performance Web-mapping components and controls for embedding maps in ASP.NET applications (Web forms). The Smart Eco-path Finder can be reached at the following URL: http://land.geo.tsukuba. ac.jp/ecowalker/ecowalker_eng.aspx.

GRAPHICAL USER INTERFACE

Under the Smart Eco-path Finder, users can select a walking target in three ways, selecting location by address, selecting location by public facility names, and selecting X, Y coordinates by clicking on a map. Users can plan and arrange multistop trips by adding destinations one by one and moving them up and down with the arrow keys. The user interface is designed to fit common smartphone screen resolutions (see Figure 5).

SHORTEST VERSUS GREENEST

Calculation of shortest versus greenest path is similar to short-est versus quickest path analysis in transportation planning. For shortest path identification, the program only uses the road distance, and for greenest path identification, the program uses greenness score as a weighted factor. Greenness score of each road

Figure 5. Graphical user interface of Smart Eco-path Finder
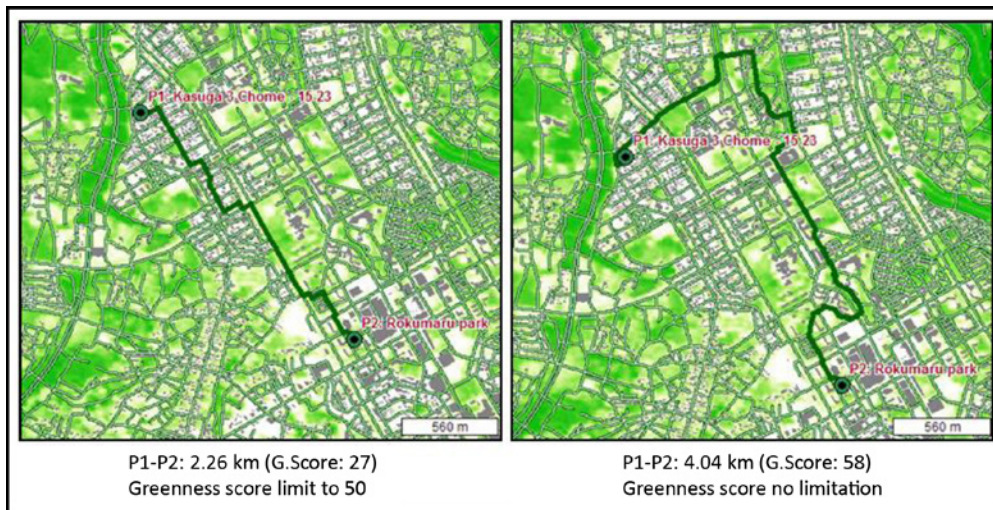
Figure 6. Greenness score limitation (threshold) for selecting between possible greenness routes
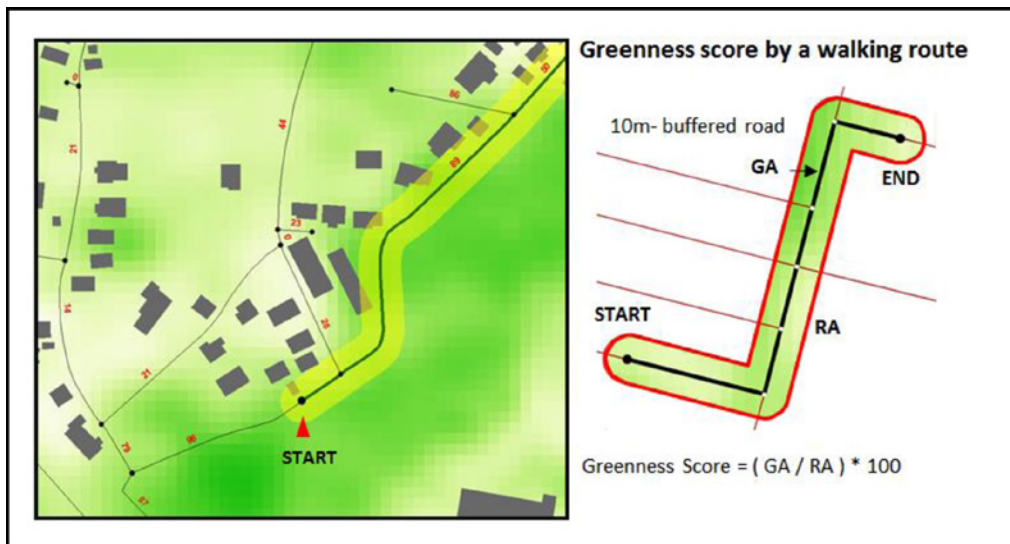
Figure 7. Calculation of greenness score for a walking route

segment was used as a weighted factor for finding the greenest path. Although, theoretically, we can get only one shortest path, there are many options to get the greenest paths. To control this, we can set up the default value while building a road network data model that is similar to speed limitation in transportation analysis. This default value limits the greenness score during the process. Here we do not limit the greenness score. Therefore, the result is the most greenness route from all possible greenness routes (shown in Figure 6). The algorithm used for finding the shortest and greenest path is modified Dijkstra's algorithm provided by VDS Technologies.

GREENNESS SCORE

The calculation of get score by walking route is based on equation (2). This measurement is ideal to inform people who want to make outdoor recreation or exercise activities a part of their daily or weekend routines. The calculation of total greenness score for one route is based on a ten-m buffered zone of the route, because start and last paths of the route could be anywhere on the street, we cannot use the score for the whole segment (see Figure 7).

$$\text{Get Score by a Walking Route} = (GA / RA) * 100 \qquad (2)$$

GA = Green Area
RA = 10 m buffered Route Area

HOW IT WORKS

Under the Smart Eco-path Finder, users are enabled to plan multistop trips by providing three selection modes, named as select by address, select by public facility, and interactive selection by clicking on the map. The route can be selected as either the shortest or greenest path (shown in Figures 8 and 9) with a single or loop
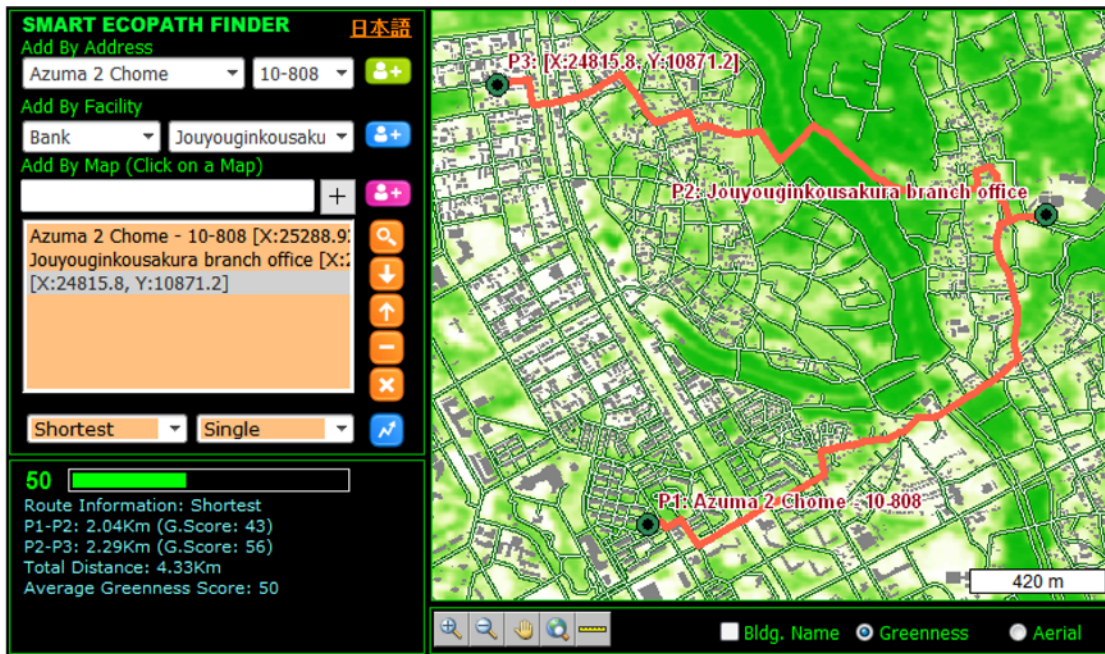
Figure 8. Shortest path using Smart Eco-path Finder (single route)
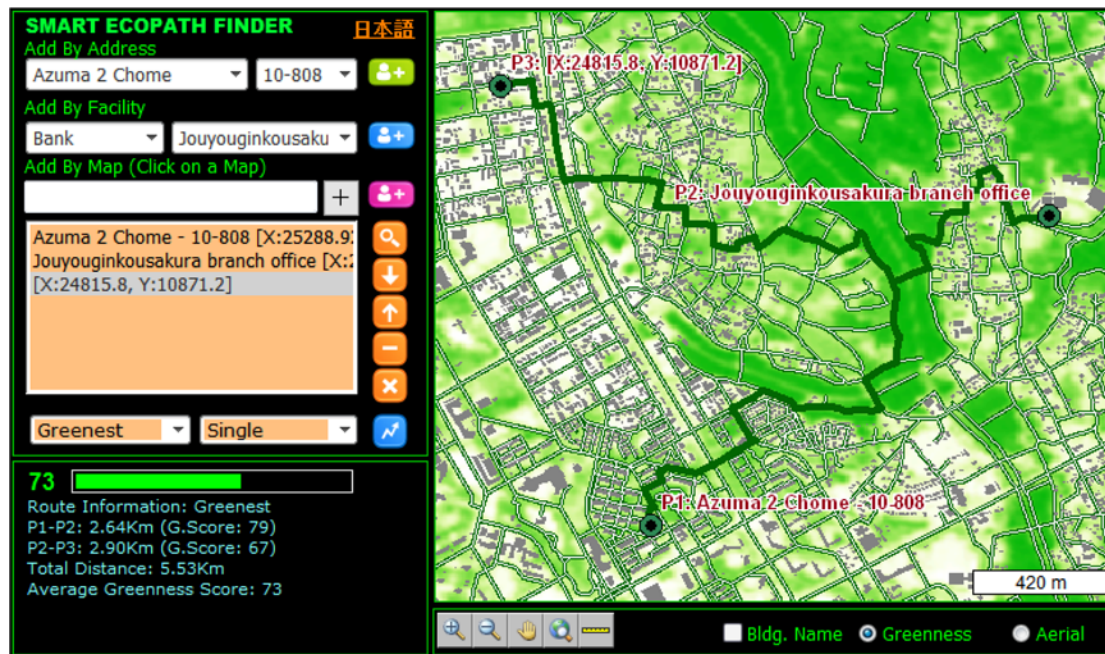
Figure 9. Greenest path using Smart Eco-path Finder (single route)
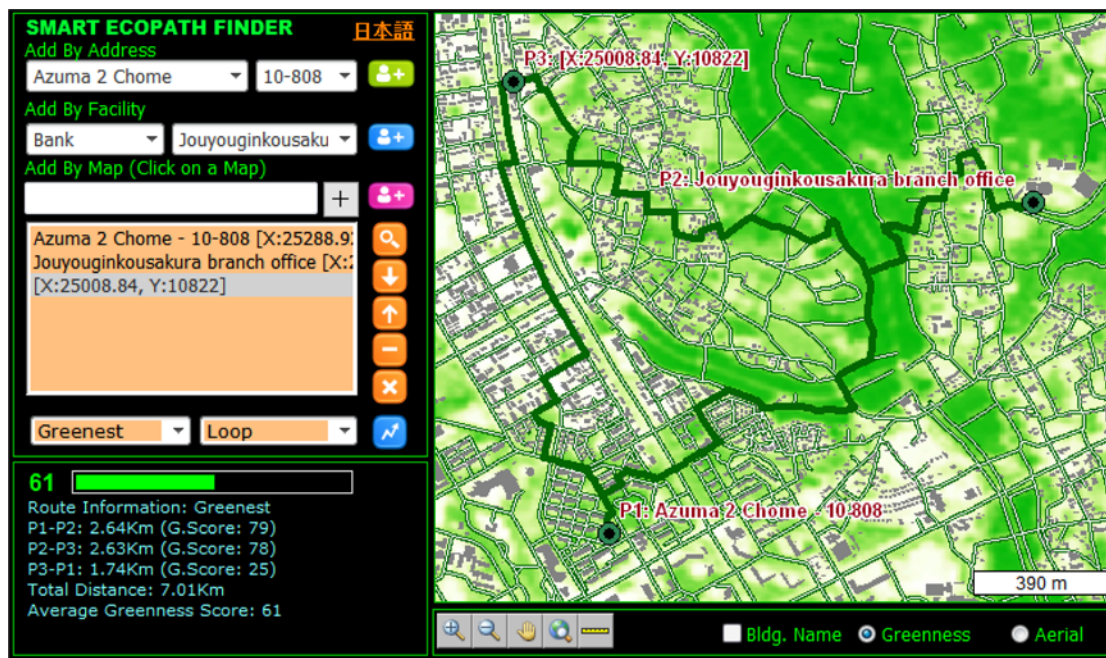
Figure 10. Greenest path using Smart Eco-path Finder (loop route)

route type. Single route type will be a multistop trip plan, and a loop route type will be for daily walking activities. For example, a user may want to walk from home to the library and continue to walk to the park and then return home (as shown in Figure 10). The user also can switch the base map to either greenness image (intensity of NDVI) or aerial image, which is suitable for real-world information (see Figure 11).

CONCLUSION

The main purpose of this paper was to introduce a Web-based interactive Eco-path Finder "Eco Smart Walker" for local residents of Tsukuba City to help with their ecobased spatial decision making and to encourage them to walk and save energy. Our work will help them make neighborhood environmental quality assessments, find the shortest or greenest path to walk to improve their physical health, and choose places to live with green spaces to improve their health and welfare. Urban green spaces are important for human mental health improvement, social cohesion, reducing the urban heat island effect, improving air quality, performing green exercises, and saving energy by reducing cooling and heating demands. GIST (GIScience and Technology) provides the identification of green spaces, calculation of greenness score, and route path analysis to improve local residents' decision making. Moreover, modern spatial Web technology (Web-GIS) is more accessible to a much wider audience than is traditional GIS. The general public now can directly access spatial information and see the results through their Web browsers without any installation of GIS software. The system itself is reusable and updatable.

About the Authors

Dr. Ko Ko Lwin is a researcher in the Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan. His research interest includes geospatial data acquisition methods and techniques, remote sensing, spatial analysis, urban information systems, and GIS application and system development.

Division of Spatial Information Science
Graduate School of Life and Environmental Sciences
University of Tsukuba
Tennodai 1-1-1, Tsukuba, Ibaraki, Japan 305-8572
E-mail: kokolwin@live.com

Dr. Yuji Murayama is a professor in the Division of Spatial Information Science, Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan. He teaches and researches issues on human geography, GIS and spatial analysis, and urban transportation geography.

Division of Spatial Information Science
Graduate School of Life and Environmental Sciences
University of Tsukuba
Tennodai 1-1-1, Tsukuba, Ibaraki, Japan 305-8572
E-mail: mura@geoenv.tsukuba.ac.jp

Figure 11. Smart Eco-path Finder with aerial image as a base map

References

Abley, S. 2005. Walkability scoping paper. Charted traffic and transportation engineering. Christchurch, New Zealand: February 2005.

Dizdaroglu, D., T. Yigitcanlar, and L. A. Dawes. 2009. Sustainable urban futures: An ecological approach to sustainable urban development. Paper read at the Proceedings of the Second Infrastructure Theme Postgraduate Conference 2009: Rethinking Sustainable Development Planning, Infrastructure Engineering, Design and Managing Urban Infrastructure.

Forsyth, A., and M. Southworth. 2008. Cities afoot—pedestrians, walkability, and urban design. Journal of Urban Design 13(1): 1-3.

Grahn, P., and U. A. Stigsdotter. 2003. Landscape planning and stress. Urban Forestry and Urban Greening 2(1): 1-18.

Kweon, B.-S., W. C. Sullivan, and A. R. Wiley. 1998. Green common spaces and the social integration of inner-city older adults. Environment and Behavior 30(6): 832-58.

Lehmann, T. M., C. Gonner, and K. Spitzer. 1999. Survey: Interpolation methods in medical image processing. IEEE Trans. Med. Imag. 18(11): 1049-75.

Lwin, K. K., and Y. Murayama. 2009. A GIS approach to estimation of building population for microspatial analysis. Transactions in GIS 13(4): 401-14.

Lwin, K. K., and Y. Murayama. 2011. Modelling of urban green space walkability: Eco-friendly walk score calculator. Computers, Environment and Urban Systems 35(5): 408-20.

Maas, J., R. A. Verheij, P. P. Groenewegen, S. De Vries, and P. Spreeuwenberg. 2006. Green space, urbanity, and health: How strong is the relation? Journal of Epidemiology and Community Health 60(7): 587-92.

Moudon, A. V., C. Lee, A. D. Cheadle, C.W. Garvin, D. B. Johnson, T. L. Schmid, et al. 2006. Operational definitions of walkable neighborhood: Theoretical and empirical insights. Journal of Physical Activity and Health 3(1): S99–117.

Pretty, J., J. Peacock, R. Hine, M. Sellens, N. South, and M. Griffin. 2007. Green exercise in the UK countryside: Effects on health and psychological well-being, and implications for policy and planning. Journal of Environmental Planning and Management 50(2): 211-31.

Reid, S. 2008. Fit for purpose: Evaluating walkability. Engineering Sustainability 161(2): 105-12.

Roehr, D., and J. Laurenz. 2008. Green surfaces in the city context. In the Proceedings of the Ecocity World Summit, April 21 to 26, 2008, San Francisco.

A Qualitative Framework for Evaluating Participation on the Geoweb

Blake Byron Walker and Claus Rinner

Abstract: The participatory Geoweb emerges from the synthesis of map-based online applications and Web 2.0 concepts such as user-generated content, enhanced interactivity, and cloud computing. The result is a wide range of tools and projects using these tools to communicate, collaborate, deliberate, and inform spatial decision making. This article draws upon the literature in participatory geographic information systems to propose the "3E Framework," which provides both a structured conceptual model and a practical tool for the evaluation of projects on the participatory Geoweb. The framework deconstructs participation on the Geoweb into the provider and public realms and represents the engagement, empowerment, and enactment processes. It includes 20 evaluation questions that are derived from themes in the literature.

INTRODUCTION

The Geospatial Web (Geoweb) includes a variety of interactive online mapping applications to which users can contribute contents (Scharl and Tochtermann 2007). While researchers have yet to settle on a single definition of the Geoweb, Crampton (2009) provides several common denominators of this "explosion of new spatial media on the web": cartography, citizen-orientation, and strong ties to the notion of public production of knowledge in participatory contexts ("crowdsourcing," p. 91). The term *participatory Geoweb* was coined by Sieber (2008) to describe "the involvement of advocacy nonprofits and marginalized communities . . . in the geospatial technologies of Web 2.0" (p. 59). Geoweb tools can function as powerful conduits for gathering volunteered geographic information (VGI) for collaborative planning, deliberation, and argumentation (Flaxman 2010, Goodchild 2010).

Each Geoweb application has its own objectives, target users, types of contributions, spatial contexts, and other attributes, resulting in substantial difficulties when trying to establish metrics for success and effectiveness. To address this challenge, this article builds on established findings in the geographic information science literature to propose a novel framework, which is designed to evaluate participatory initiatives that utilize Geoweb technologies, and may be extensible to public participation geographic information systems (PPGIS) and other participatory media. It also provides a structured approach for contextualizing the processes inherent in user participation through these media.

The following section provides the research context, background on participation, and spatial decision making, and introduces concepts used in the proposed framework. This lays the groundwork for the subsequently presented description of the "3E Framework" with its provider and public realms, interaction space, and engagement, empowerment, and enactment processes. To conclude, we describe how this framework may be deployed, outline some of its limitations, and make a call for continued efforts to unite participatory Geoweb theories with practice.

RESEARCH CONTEXT

Origins of the Geoweb

Recent years have seen the emergence of Web 2.0 as the next generation of online networking, where users contribute content by uploading files, writing reviews, rating contents, and otherwise contributing to a "collective intelligence" (Lévy 1997, Gordon-Murname 2006). While O'Reilly (2005) is credited with coining the term *Web 2.0,* it first appears six years earlier in an article by DiNucci (1999). Controversy aside, the fact remains that the Web is shifting from a static information source to an interactive platform integrating user contributions, and efforts to leverage these capabilities to empower citizens in spatial decision-making processes abound (e.g., Ward, Gibson et al. 2003; Miller 2006; Tulloch 2007; Mericskay and Roche 2010).

As the Web continues to evolve from a one-way information conduit to a multidirectional interactive space, so, too, does it adopt new spatial contexts. The concepts underpinning the Geoweb can be traced back to Herring (1994), whose U.S. Department of Defense paper postulated the implementation of "spatial indexing geometry" for battlefield mapping and simulation. He called this the beginning of the "spatialization of the internet" (p. 1). During the following decade, interactive mapping applications exploded across the Web, backed by information technology giants Google, Microsoft, and Yahoo (Goodchild 2005 Craglia, Goodchild et al. 2008). It did not take long for users worldwide to recognize the value of interactive "cybercartography"; Peterson's (2005) study concluded that maps were, at the time, the second most frequent request over the Internet (next to weather forecasts) (Taylor and Claquard 2006).

The synthesis of Web 2.0 concepts with online mapping technologies produces the participatory Geoweb, a "phenomenon that has taken the world of geographic information by storm" (Goodchild 2007, Maguire 2007, Tulloch 2007). In the same

fashion that Web 2.0 encourages crowdsourcing (the accumulation of data submitted by many users, e.g., Wikipedia), Goodchild (2007, 2008) brought attention to the notion of volunteered geographic information (VGI), describing user-generated, spatially referenced information (e.g., Wikimapia). In this work, Goodchild conceptualizes "citizens as sensors" for the gathering of intelligence directly from sources in the field. In recent years, the concept of the *produser* has been developed to describe a user of Web 2.0 technologies who both accesses and contributes information (Budhathoki et al. 2008, Coleman et al. 2009).

However, this raises concerns among other researchers, who identify abundant questions and concerns with the quality and veracity of VGI. Research in PPGIS had been active before the concept of VGI was formally established, and a body of literature tying theory with the exponentially growing use of geographic tools in participatory decision making continues to mature (Pickles 1997; Kingston, Carver et al. 2000; Jankowski and Nyerges 2001; Breitbart 2003; McCall 2003; Sieber 2006; Pain and Kindon 2007).

PERSPECTIVES ON PPGIS

Many participatory Geoweb applications can be conceptualized in PPGIS terms, for they are tools for implementing spatial information and gathering input from the public. It is thus advantageous to draw upon the substantial literature examining the use of PPGIS to engage stakeholders, gather information, and inform decision makers. This body of research ranges from applied studies of PPGIS in real-world projects (Han and Peng 2003) to theoretical explorations of these technologies from a variety of perspectives (Bussi 2001, Ghose 2001, Kwan 2002, Sieber 2006). Such previous efforts to conceptually synthesize GIS and public participation provide fundamental building blocks for the framework proposed here.

The features by which the participatory Geoweb is most commonly distinguished from PPGIS are scale (of both number of participants and volume of data), integration of increasingly ubiquitous mobile and social networking technologies, decreasing levels of analytical capability, and less required technological expertise for use (Cinnamon and Schuurman 2010, Crampton et al. 2013).

Peng (1999, 2001) provides frameworks for the planning of Internet-based GIS systems from an IT perspective, and also proposes a taxonomy for Web-based public participation systems, based on their functions and contents. While this matrix-based classification system provides a straightforward typology for characterizing Web-based participation tools, it does not delve into the abstract social and political processes underlying participation. Carver et al. (2001) discuss their findings from two PPGIS case studies and forecast the expansion of Internet-based public participation. Many of the issues they raise continue to influence participation on the Geoweb, namely access to the Internet, IT knowledge, and political factors affecting the implementation of publicly generated information, concerns that are further explored by McCall (2003). To address these aspects, Jankowski and Ny-

erges (2003) propose the EAST 2 framework for examining the interaction of sociopolitical constructs and influences that occur with the use of participatory GIS. Dragicevic and Balram (2004) provide a framework for conducting Web-based, collaborative spatial deliberation in collaborative resource management, with a focus on equity and access. Sieber (2006) identifies four broad social themes found throughout the PPGIS literature, examines specific elements of these themes and their interactions, and proposes a framework for academic evaluation of PPGIS along these lines. A common theme throughout the PPGIS literature is the effort to reconcile and position GIS concepts within the broader participation literature, generally drawn in sociopolitical motifs.

SOCIOPOLITICAL DIMENSIONS OF PARTICIPATION

Participation in the decision-making context does not settle on a singular definition, although many authors do point to its roots in democratic theory and empowerment of the public (Rosenstone and Hansen 1993; Perkins, Brown et al. 1996; Bussi 2001; Beierle and Cayford 2002; Abelson, Forest et al. 2003; Nyerges 2005; Miller 2006). Wang and French (2008) divide the ingredients of e-participation into methodology, community, and technology, and proceed to dissect "common understanding" into actionable elements of an "e-democracy"; Nyerges (2005) contrasts "deliberative" and "representative" democracies, highlighting the former as relevant to participatory decision making, while the latter better describes elected representation. Schlossburg and Shuford (2005) divide participation into two base concepts: the acts of participation themselves and the broad goals of participation. The prevailing view of the latter has been that of a catalyst for the shift of power from traditional hierarchical structures to the grassroots, originating from Arnstein (1969). She defined participation as "the redistribution of power that enables the have-not citizens, presently excluded from the political and economic processes, to be deliberately included in the future" (p. 351). Arnstein's "ladder of participation" exemplifies this dynamic along a scale from "manipulation" to "citizen control." Conversely, Wiedemann and Femers (1993) model public participation from a government perspective, characterizing levels of participation by their involvement in a decision-making process. It is, however, necessary to heed the differentiation between the roles of the leaders and the public, as Connor (2007) explains, taking the view that participation is a tool for the prevention and mitigation of conflict between these two bodies.

ACTS OF PARTICIPATION

The instantiations of participation in spatial decision making vary widely, PPGIS being the dominant medium established in the literature. However, PPGIS represent only one genus of a rapidly evolving mass of geographic tools for public use. The focus of this work is the participatory Geoweb species specifically tasked with collecting public knowledge to inform spatial decision making.

The acts of participation are examined by Nyerges (2005),

who identifies four independent types of participatory actions: data operation, analysis, speech/dialogue, and deliberation, while Rinner et al. (2011) discuss the "analytic-deliberative" perspective of decision making, identifying information processing and dialogue as respectively subordinate stages to analysis and deliberation.

The participation acts and overarching goals can be reconciled by implementing Jankowski and Nyerges' (2001) "macro-micro approach" to structuring decision situations. This system divides each macro phase (intelligence, design, and choice) into four microactivities (gathering, organization, selection, and review), simplifying the analytic-deliberative decision-making process. Their EAST2 framework is used to interpret the macrophases, and is accompanied by eight constructs in decision-making situations that seek to explain the different perspectives involved. Future investigation of the acts of participation need not be limited to a geographical realm, for example, Brandt (2006) investigates exploratory design games as potential participation frameworks.

While previous research has been successful in theoretically and empirically examining the social, political, and material dimensions of public participation in spatial decision making, no distinct method for implementing these findings into the planning or evaluation phases of a project is yet found in the literature. To address this gap, we propose a novel framework accompanied with specific questions for evaluating projects on the participatory Geoweb.

THE 3E FRAMEWORK

The 3E Framework links theory and practice by using the aforementioned dimensions of participation and the Geoweb to guide the post hoc evaluation of projects that use the participatory Geoweb to inform spatial decision making. As illustrated in Figure 1, this framework deconstructs participation into a three-stage process (engagement, empowerment, enactment) occurring across three conceptual spaces (provider realm, public realm, Geoweb interaction space). Twenty qualitative questions pertaining to these spaces and processes are designed as guideposts to assist with the evaluation of a project on the participatory Geoweb. In this section, we define the framework elements and present the accompanying questions.

KEY CONCEPTS

Many authors point to *engagement* as central to democracy (Owens 2000, Williams 2004, Rinner and Bird 2009, Boulton 2010). Engagement in this framework encompasses both the act of securing a space in which project entities interact with participants ("publics"), and the actors and networks within that space itself (Bachimont 2000, Rowe and Frewer 2000, Gagnon and Fortin 2002, Ghose 2007). From the results of a Canada-wide survey, Robinson and Gore (2005) identify lack of public engagement as the primary limiting factor in the development of municipal policies regarding climate change. To better reach out to people, "the public" must be defined and identified. Schlossburg and

Shuford (2005) propose three definitions of who constitutes "public" in public participation: "those affected by a decision or programme"; "those who can bring important knowledge or information to a decision or programme"; and "those who have power to influence and/or affect implementation of a decision or programme" (p. 18). The publics are increasingly seen as "produsers" of VGI (Budhathoki et al. 2008, Coleman et al. 2009). But this conceptualization of publics in the participatory context relies on notions of representation and expertise, which can be difficult to evaluate (Rowe and Frewer 2000, Barnes, Newman et al. 2003). Furthermore, map literacy and access (digital inequalities) are significant factors in public engagement (Crampton 2009, Haklay 2013). Creighton (1983) describes five criteria for characterizing publics in participation processes: proximity to the space in question; economic stakes; utilization of the space in question; social stakes; and value systems. These are used to plan and evaluate the public realm and the engagement phase of the 3E Framework, as described below.

The *empowerment* phase in this framework is predicated on the view that participation is a means to politically promote citizen interests and flatten hierarchies (Arnstein 1969; Craig, Harris et al. 2002; McCall 2003). While Haklay (2013) argues that the Geoweb and other technological means of participation serve an elite body of users, resulting in a frequent overstatement of the Geoweb's ability to depoliticize decision making, others argue that the Web has the potential to empower publics more than ever by providing nearly infinite information at one's fingertips (Sieber 2006, Tulloch 2007). The proposed 3E Framework directs project planners and/or evaluators to examine power dynamics, hierarchies, and social structures inherent in participation, predicated by Ghose's (2007) assertion that the context of a specific participatory project space embodies hierarchies while providing opportunities for participants to transcend traditional power networks. This includes consideration of the "information needs" of publics to participate in a meaningful way (Jankowski, Nyerges et al. 2006). Also addressed here is Nyerges' (2005) assertion that "access to voice" is the critical basis on which participation occurs. Perkins and Brown (1996) utilize an ecological framework to predict public participation in community groups. Their physical, economic, and social indicators are reflected in the empowerment phase in the 3E Framework.

Enactment as the third process in the framework refers to the implementation of the results from participation into a decision-making process. While some authors argue that participation itself can be more significant than its formal outcomes (e.g., in fostering community spirit), the ultimate goal of participation is to include publics in the decision-making process (Arnstein 1969, Schlossburg and Shuford 2005, Miller 2006, Rambaldi 2006). The enactment phase examines the ways in which input gathered through participation is considered in the decision-making process. However, this is not always straightforward. For example, McCall (2003) identifies implementation issues with indigenous spatial knowledge. Boulton (2010) summarizes the concern that users must see results if they are to continue to
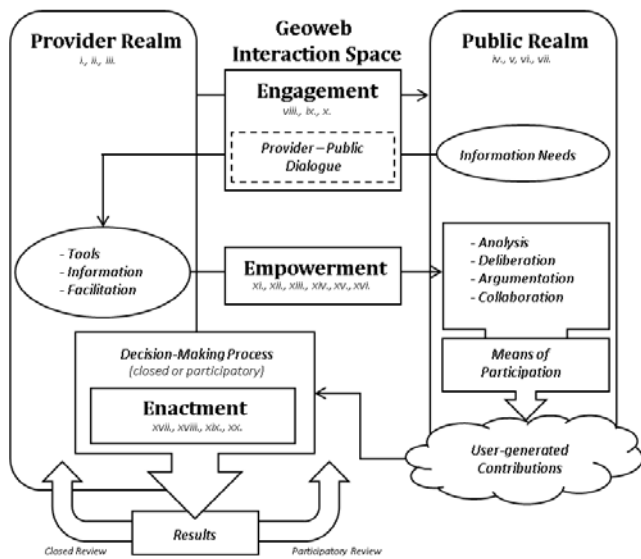
Figure 1. The 3E Framework consisting of provider and public realms, interaction space, and engagement, empowerment, and enactment processes

Figure 2. Evaluation questions derived from the 3E Framework

embrace the participatory Geoweb, and Bussi (2001) argues that there exists a fundamental gap between geography and democracy, offering six hypotheses for its cause. He points toward local and small-scale development as the first steps toward a reconciliation of the two. Elwood (2006) discusses gaps in knowledge production and subsequent implementation in decision making, a problem that the 3E Framework does not propose to overcome; however, project planners/evaluators using the framework are directed to specifically consider this phase.

To deploy this framework for the evaluation of a project on the participatory Geoweb, we provide 20 questions pertaining to various components of the framework (numbered in Figure 2). These are designed to assist the evaluator in teasing out the various dimensions of participation on the Geoweb as they apply to the project being considered.

THE PROVIDER REALM

The provider realm encompasses the project administrator's arena of operations, including project design, tool development, and (often) decision making. Three questions are used to evaluate the provider realm:

i.   **Who is the project provider?**
     Although Geoweb projects may not always follow a clear provider-user dichotomy, we argue that a provider, facilitator, or initiator must exist for a project to emerge. For example, Twitter as a generic platform can act as facilitator for a set of related tweets. If a Twitter user, or group of users, start employing a new hashtag and encourage others to join a discussion under this hashtag, they can be considered initiators of a project. Finally, if an organization invites comments via Twitter under a given hashtag on a given

topic, we also would consider them under the provider realm although they would not provide the project infrastructure. Therefore, this first evaluation question asks to explicitly identify who is involved in the provider realm, a necessary step to later analyze the power dynamics between provider and public. In this case, "who" refers to the people or organizations initiating, facilitating, or administrating the project. Often, the distinction between providers and the public may be blurred. Particular attention should be paid to the actors' authority and expertise.

ii. **Why is the project being conducted?**
     The overarching project goals are stated, including design objectives and specific delineation of the space in question. Identifying the wider project goals and scope is useful for analyzing the methods of engagement and participation, and determining if these conjugate.

iii. **Why is public participation sought?**
     Recognizing that public participation is only one means of gathering information and opinions to inform spatial decision making, it is important to make clear the reasons why public input was required or desired. What knowledge or power is held by the public that makes their participation valuable?

PUBLIC REALM

The public realm is the conceptual space where participants analyze, deliberate, and collaborate to produce participatory contributions. Participatory contributions include input in the form of comments, data submissions, and media uploading, and is also known as "user-generated content" or VGI in geospatial applications. In this context, "public" refers to any target user group, not exclusively a civilian body of the citizenry (Rowe and Frewer 2000).

iv. **Who comprise the target publics?**

Target publics can include individuals, community groups, or highly structured organizations. The physical and social settings in which participants are situated could have a significant impact on their contributions, including demographic characteristics (Cinderby 2010).

v. **What are the publics' spatial relationships to the project objective?**

Vital to assessing the contributions of a public is an understanding of the public's use of the space in question. This includes tangible and symbolic uses. The place of residence and, to a lesser degree, work are key components.

vi. **What other motivations for participation may the publics have?**

Participants may be influenced by affiliations, preferences, beliefs, and other motivators, affecting the content and quality of their participatory contributions (Coleman et al. 2009). Creighton's (1983) five criteria for evaluating publics provide a guide for identifying such influences: proximity to the space in question; economic stakes; utilization of the space in question; social stakes; and value systems. Using these criteria, we can attempt to estimate the nature and strength of such influences on the content of contributions. For example, a project to collect spatially referenced citizen input on an urban development is likely to gather polarized opinions; recognizing the motivations of the participating publics will allow evaluators to determine whether the participatory design encourages bias or division of opinion.

vii. **What is the nature of the target publics' expertise in the matter?**

This question draws attention to what the participants know and how their input is assessed for credibility. By examining the knowledge gap between provider and public, the project evaluator can determine if the participants are, in fact, the right group to consult.

ENGAGEMENT

The first phase of participation involves selecting and contacting the target users (publics). The conceptual space between provider and public realms in the 3E Framework is termed the "interaction space," through which engagement occurs. While serving as the connecting medium between the two realms, this element also represents other dynamics between the provider and publics. To determine how the relation between these two entities affects participation and the project as a whole, the question is posed:

viii. **What are the means of communication between provider and target publics?**

The methods by which publics are recruited into the participation process are critical to engaging the targeted group. For example, while mobile services are an excellent way to contact youth, this may not be the best method to get contributions from seniors, an instance where in-person facilitation may prove more effective. This question

reflects the simple requirement that a participatory project on the Geoweb can be effective only if the target publics are reached. Furthermore, the methods of communication should facilitate a two-way dialogue in keeping with Wang and French's (2008) "community" criterion for e-democracy.

ix. **What is the desired number of participants and frequency of participation?**

While Geoweb technologies are able to efficiently handle larger numbers of participants and contributions relative to traditional methods, an approximation of the number of expected contributions ensures that handling the received data is within the capabilities of the project provider. While some projects only require one-time contributions, others rely on sustained participation over time and throughout the entire spatial decision-making process. If sustained participation is desired, whether or not this is achieved can help to identify areas for improving engagement.

x. **Why is the Geoweb a desired medium for participation?**

This question directs evaluators to consider why other media were not deployed in place of interactive Web mapping. Themes here may include the spatial nature of the knowledge being sought, the Web presence of the target publics, and any barriers posed by other options (e.g., focus groups, telephone surveys). Findings here may reveal effective alternatives or complementary approaches to the Geoweb component of a project.

EMPOWERMENT

The second phase of participation, empowerment, addresses to participants' information needs, computer literacy, knowledge, and stakeholder status (with respect to sociopolitical dynamics) (Nyerges 2005; Jankowski, Nyerges et al. 2006; Ramsey and Wilson 2006). The following evaluation questions are designed to examine several key facets of empowerment in the participatory Geoweb context.

xi. **What are the publics' information needs (i.e., what must they know to participate)?**

Often, the core purpose of leveraging participatory technologies on the Geoweb is to collect VGI (Sieber 2006, Goodchild 2008). To effectively participate, the target users must have both the knowledge to contribute and the means of contributing that knowledge. This includes the ability to operate a participatory tool, analyze information, and, in some cases, deliberate, argue, and/or collaborate. This also could include knowledge about design practice, code, and economic factors. Furthermore, reciprocal learning should be considered, where users learn about the geographical, social, and political contexts through the process of participation, as observed by Elwood (2009).

xii. **What social, economic, and cultural requisites for participation exist?**

This question addresses issues such as Internet access, computer literacy, and language abilities. It is important to

assess who is included, but also who is excluded by a project design, to assess its effectiveness and alternative modes of action. More comprehensive guidelines for evaluating these dimensions are found in Sieber (2006) and Haklay (2013).

xiii. **What types of participatory contributions are expected?**
When considering the spatial information project providers seek, it is necessary to differentiate between observations, opinion, and contributed designs. The method of participation is critical in receiving participatory contributions of the desired type (Rinner et al. 2011). This influences the format of contributions, assessment of their validity, and how they are utilized in a decision-making regimen. This evaluation question also asks about the possibility of dialogue between participants, e.g., in a threaded discussion forum or map-based deliberation platform (Rinner et al. 2008).

xiv. **What are the spatial features of user contributions (e.g., points, polygons)?**
For example, users may create points to which they attach comments. Is the spatial representation appropriate for the phenomena being considered? Perhaps alternative spatial representations would prove more appropriate (e.g., a raster-based "paintbrush' tool to allow participants to depict land cover or a Google SketchUp-based plug-in).

xv. **What are the non-spatial features of user contributions (e.g., text, multimedia)?**
This question focuses more specifically on the thematic content of participant contributions. Examples here include stories/narratives, photographs, measurements, or polls/surveys. Often the knowledge or opinion sought from participants can be enriched with other material. This question directs the evaluator to consider whether the form of contributions was suitable for effectively gathering the desired information.

xvi. **How does participation alter the power dynamic between provider and public?**
This question draws on responses from the engagement phase, encouraging evaluators to critically examine how the provider and public interrelate, with respect to authority and common goals. Power structures and the desire for conflict or cooperation can alter participatory contributions, and an understanding of this dynamic is necessary to evaluate the participatory project as a whole. Drawing from the perspectives in the literature that participation is a fundamentally democratic method, it is necessary to consider the ramifications of the project being analyzed in these terms. Does participation have the propensity to alter power structures? While the answer remains highly subjective and relies on how contributions are utilized, it can help to identify methodological weakness in the project design. It also must be considered that the overarching political context of a case study is vital to its interpretation (Arnstein 1969, Bussi 2001). Often, a project's effect on power dynamics

is determined by how the participatory contributions are incorporated in a decision-making process, and so this step is considered independently in the following section.

ENACTMENT

This construct poses questions about the ways in which user input gathered through the participatory Geoweb is utilized in the actual decision-making process on the provider side.

xvii. **How are contributions implemented into the decision-making process?**
Considering how participatory contributions are utilized in decision making is useful in determining whether the method of participation and format of input is appropriate. The project administrators may directly handle contributions, or they may be processed and passed on to a decision-making body.

xviii. **How do the results correspond with the publics' motivations for participation?**
Reflecting back on questions v and vi, this question directs the project evaluator to examine the differences between the project outcomes and the participants' motivations for involvement. This is important for determining whether public participation effectively influenced project outcomes and provides insight into the publics' sense of engagement with the decision-making process (Elwood 2009).

xix. **Are the results available to the public?**
This question refers to the transparency of the decision-making process, whether the participants are aware of how their contributions were used and to what ends. Informing participating publics of the role their contributions make in a project may prove useful for sustained engagement and a sense of empowerment, along with an opportunity for critical reflection on their collaborative role in spatial decision making (Elwood 2009).

xx. **What, if any, is the review process, and is it also participatory?**
While the 3E Framework itself can be used for postproject review, this question seeks to identify any other review measures taken and whether or not project contributors also participate in the review. Perhaps those who contributed are able to suggest improvements to the Geoweb tools and procedures, or propose novel ideas for gathering, processing, and implementing contributions.

DISCUSSION AND CONCLUSION

In the post hoc evaluation of a participatory Geoweb project, the 3E Framework can be deployed to structure the project review. The evaluation questions are mapped to specific elements of the framework, and exploring each of these will assist project evaluators in relating literature-derived features of effective participation to actual project results. This work represents a step toward reconciling PPGIS theory with practical applications of the Geoweb, but also may have value in conceptualizing, framing,

and evaluating other participatory and VGI initiatives.

The 3E Framework also may assist in the planning of future projects on the participatory Geoweb. For this purpose, the evaluation questions would be answered with intended or expected project characteristics. Discrepancies between these and factual information about the provider or public realm, and the engagement, empowerment, or enactment processes being planned, could trigger adjustments to the project prior to, and in the process of, its deployment.

Several aspects of the framework require additional study. For one, the nature of this approach suggests an opportunity for participatory review, a crucial stage often overlooked in the literature. The framework questions may represent a bias toward the provider, as participating publics may not be aware of a project's goals and means of enactment, and may have conflicting views on associated power dynamics. As such, collaboration with the target publics during the review process is encouraged. While the division of a participatory project into three conceptual spaces is convenient, the spaces remain strongly connected along many lines, and the framework may oversimplify large-scale participatory projects and complex methods of participation. As such, context-specific questions could provide additional insight into the realms, actors, and processes behind a participatory initiative.

Scalability of Geoweb projects is of significant concern and should be addressed in future research. Some authors point to localized, small-scale participation as more fundamentally democratic (Arnstein 1969, Bussi 2001), while others highlight technical approaches to address scalability (Sani and Rinner 2011). To implement more quantitative means of project evaluation, a scoring system could be devised with the addition of ordinal "provider satisfaction" and "public satisfaction" columns to the questions matrix. The responses and the perceived strengths of each project element then would be rated by all participants and standardized to produce a score. In this way, the effectiveness of a project across the framework could be rated and compared to other such projects, lending an additional dimension to the analysis.

As the Geoweb continues to grow, so, too, does the body of questions surrounding its applications to participatory design, planning, and policy making. On the one hand, the increasing complexity and sophistication of information technology may create problems for understanding Geoweb tools and their contexts. On the other hand, Geoweb tools are becoming increasingly streamlined and easy to use, thus potentially hiding the complexity of decision problems presented through them. This research offers a deconstructive method for analyzing the processes on the participatory Geoweb to better address questions surrounding its ability to democratize societal decision making, and may be extensible to PPGIS and other participatory media. An understanding of the positives and negatives of utilizing Web technologies to foster and implement public input in decisions traditionally held within rigid power structures and hierarchies benefits our transitioning views of democracy in the 21st century, as the Web 2.0 continues to expand its influence on the people and processes that design our spaces.

Acknowledgments

About the Authors

Blake Byron Walker is a Ph.D. student in the department of geography at Simon Fraser University (SFU), Vancouver, Canada, specializing in spatial epidemiology and GIS for health informatics. Walker completed the present research as a student in the BA in Geographic Analysis program at Ryerson University. His Master's research at SFU examined geographical patterns of violent trauma, pedestrian injury, and graffiti, while his doctoral research focuses on the spatial-temporal trends in head and neck cancers in British Columbia.

Department of Geography
Simon Fraser University
8888 University Drive
Burnaby, British Columbia V5A 1S6
Canada
Phone: (778) 782-4987
E-mail: bwalker@sfu.ca

Claus Rinner is an associate professor in the department of geography at Ryerson University, Toronto, Canada. Rinner holds degrees in applied systems sciences, applied mathematics and social sciences, and geography from the Universities of Montpellier, France, and Osnabrück and Bonn, Germany. Within geographic information science, his research contributes to enhance the mapping and analysis functions in GIS to support visual thinking and to facilitate effective and sustainable spatial decision making.

Department of Geography
Ryerson University
350 Victoria Street
Toronto, Ontario M5B 2K3
Canada
Phone: (416) 979-5000, x2686
E-mail: crinner@ryerson.ca

References

Abelson, J., P. G. Forest, J. Eyles, P. Smith, E. Martin, and F. P. Gauvin. 2003. Deliberations about deliberative methods: Issues in the design and evaluation of public participation processes. Social Science and Medicine 57(2): 239-51.

Arnstein, S. 1969. A ladder of citizen participation. Journal of the American Institute of Planners 35(4): 216-24.

Bachimont, B. 2000. Engagement sémantique et engagement ontologique: Conception et realisation d'ontologies en Ingénierie des connaissances. In Charlet, J., M. Zacklad, G. Kassel, and D. Bourigault, Eds., Ingénierie des connaissances, évolutions récentes et nouveaux défis. Paris: Eyrolles.

Barnes, M., J. Newman, A. Knops, and H. Sullivan. 2003. Constituting the "public" in public participation. Public Administration 82(1): 379-99.

Beierle, T. C., and J. Cayford. 2002. Democracy in practice: Public participation in environmental decisions. Washington, DC: Resources for the Future.

Boulton, A. 2010. Just maps: Google's democratic map-making community? Cartographica, 45(1): 1-4.

Brandt, E. 2006. Designing exploratory design games: A framework for participation in participatory design? Proceedings of the 9th Participatory Design Conference, August 1 to 5, 2006, Trento, Italy.

Breitbart, M. M. 2003. Participatory research. In Clifford. N., and G. Valentine, Eds., Key methods in geography. London: Sage.

Budhathoki, R., B. Bruce, and Z. Nedovic-Budic. 200. Reconceptualizing the role of the user of spatial data infrastructure. Geojournal 72(3/4): 149-60.

Bussi, M. 2001. Géographie, démocratie, participation: Explication d'une distance, arguments pour un rapprochement. Géocarrefour 76(3): 265-72.

Carver, S., A. Evans, R. Kingston, and I. Turton. 2001. Public participation, GIS, and cyberdemocracy: Evaluating online spatial decision support systems. Environment and Planning B: Planning and Design 28(6): 907-21.

Cinderby, S. 2010. How to reach the "hard-to-reach": The development of participatory geographic information systems (P-GIS) for inclusive urban design in UK cities. Area 42(2): 239-51.

Cinnamon, J., and N. Schuurman. 2010. Injury surveillance in low-resource settings using geospatial and social Web technologies. International Journal of Health Geographics 9(25). Doi: 10.1186/1476-072X-9-25.

Coleman, D. J., Y. Georgiadou, and J. Labonte. 2009. Volunteered geographic information: The nature and motivation of produsers. International Journal of Spatial Data Infrastructures 4: 332-58.

Connor, D. M. 2007. A new ladder of citizen participation. National Civic Review 77(3): 249-57.

Craglia, M., M. F. Goodchild, A. Annoni, G. Camara, M. Gould, W. Kuhn, and E. Parsons. 2008. Next-generation digital earth. International Journal of Spatial Data Infrastructure Research 3: 146-67.

Craig, W., T. M. Harris, and D. Weiner, Eds. 2002. Community participation and geographic information systems. London: Taylor and Francis.

Crampton, J. W. 2009. Cartography: Maps 2.0. Progress in Human Geography 33(1): 91-100.

Crampton, J. W., M. Graham, A. Poorthuis, T. Shelton, M. Stephens, M. W. Wilson, and M. Zook. 2013. Beyond the geotag: Situating "big data" and leveraging the potential of the Geoweb. Cartography and Geographic Information Science 40(2): 130-9.

Creighton, J. L. 1983. Identifying publics/staff identification techniques. Fort Belvoir: The Institute for Water Resources.

DiNucci, D. 1999. Fragmented future. Print 53(4): 32.

Dragicevic, S., and S. Balram. 2004. A Web GIS collaborative framework to structure and manage distributed planning processes. Journal of Geographical Systems 6(2): 133-53.

Elwood, S. 2006. Negotiating knowledge production: The everyday inclusions, exclusions, and contradictions of participatory GIS research. The Professional Geographer 58(2): 197-208.

Elwood, S. 2009. Integrating participatory action research and GIS education: Negotiating methodologies, politics, and technologies. Journal of Geography in Higher Education 33(1): 51-65.

Flaxman, M. 2010. GeoDesign: Fundamental principles. Paper presented at the 2010 GeoDesign Summit, January 6 to 8, 2010, Redlands.

Gagnon, É., and A. Fortin. 2002. L'espace et le temps de l'engagement bénévole: Essai de définition. Nouvelles Pratiques Sociales 15(2): 66-76.

Gallis, M. 2010. GeoWeb 2.0. Paper presented at the 2010 GeoDesign Summit, January 6 to 8, 2010, Redlands.

Ghose, R. 2001. Use of information technology for community empowerment: Transforming geographic information systems into community information systems. Transactions in GIS 5(2): 141-63.

Ghose, R. 2007. Politics of scale and networks of association in public participation GIS. Environment and Planning A 39(8): 1,961-80.

Goodchild, M. F. 2005. What does Google Earth mean for the spatial sciences? Paper presented at the Biennial Conference of the Spatial Sciences Institute, September 12 to 16, 2005, Melbourne, Australia.

Goodchild, M. F. 2007a. Citizens as sensors: The world of volunteered geography. GeoJournal 69(4): 211-21.

Goodchild, M. F. 2007b. Citizens as sensors: Web 2.0 and the volunteering of geographic information. Geofocus 7: 8-10.

Goodchild, M. F. 2008. Assertion and authority: The science of user-generated geographic content (proceedings of the colloquium for Andrew U. Frank's 60th birthday). Paper presented at the Geoinfo 39, Vienna, Austria.

Goodchild, M. F. 2010. Spatial by design. Paper presented at the 2010 GeoDesign Summit, January 6 to 8, 2010, Redlands.

Gordon-Murname, L. 2006. Social bookmarking, folksonomies, and Web 2.0 tools. Searcher: the Magazine for Database Professionals 14(6): 26-38.

Haklay, M. 2013. Neogeography and the delusion of democratization. Environment and Planning A 45(1): 55-69.

Han, S. S., and Z. Peng. 2003. Public participation GIS (PPGIS) for town council management in Singaport. Environment and Planning B: Planning and Design 30: 89-111.

Herring, C. 1994. An architecture of cyberspace: Spatialization of the Internet. Champaign: U.S. Department of Defense.

Jankowski, P., and T. Nyerges. 2001. Geographic information systems for group decision making. London: Taylor and Francis.

Jankowski, P., and T. Nyerges. 2003. Toward a framework for research on geographic information-supported participatory decision making. URISA Journal 15(APA I): 9-17.

Jankowski, P., T. Nyerges, S. Robischon, K. Ramsey, and D. Tuthill. 2006. Design considerations and evaluation of a collaborative, spatio-temporal decision support system. Transactions in GIS 10(3): 335-54.

Kingston, R., S. Carver, A. Evans, and I. Turton. 2000. Web-based public participation geographical information systems: An aid to local environmental decision making. Computers, Environment and Urban Systems 24(2): 109-25.

Kwan, M. P. 2002. Feminist visualization: Re-envisioning GIS as a method in feminist geographic research. Annals of the Association of American Geographers 92(4): 645-61.

Lévy, P. 1997. Collective intelligence: Mankind's emerging world in cyberspace. Cambridge: Perseus.

Maguire, D. J. 2007. GeoWeb 2.0 and volunteered GI. Paper presented at the Workshop presented on Volunteered Geographic Information, Santa Barbara. Http://www.ncgia.ucsb.edu/ projects/vgi/docs/position/Maguire_paper.pdf.

McCall, M. 2003. Seeking good governance in participatory GIS: A review of processes and governance dimensions in applying GIS to participatory spatial planning. Habitat International 27(4): 549-73.

Mericskay, B., and S. Roche. 2010. La néogéographie ou la cartographie à l'ère du Web 2.0. Géoinfo: chroniques du Québec géographie, dated July 2010. Retrieved from http://www.quebecgeographique.gouv.qc.ca/approfondir/bibliotheque/geoinfo/geoinfo-juillet-2010.asp.

Miller, C. 2006. A beast in the field: The Google Maps mashup as GIS/2. Cartographica 41(3): 187-99.

Nyerges, T. 2005a. Public participation GIS support for transportation improvement decision making in the age of instant access. Paper presented at the Symposium on Societies and Cities in the Age of Instant Access, Salt Lake City.

Nyerges, T. 2005b. Scaling-up and a grand challenge for public participation GIS. Directions Magazine 2005-09-20.

O'Reilly, T. 2005. What is Web 2.0—design patterns and business models for the next generation of software. San Francisco: O'Reilly.

Owens, S. 2000. "Engaging the public": Information and deliberation in environmental policy. Environment and Planning A 32(7): 1,141-48.

Pain, R., and S. Kindon. 2007. Participatory geographies. Environment and Planning A 39(12): 2.807-12.

Peng, Z. 1999. An assessment framework for the development of Internet GIS. Environment and Planning B: Planning and Design 26: 117-32.

Peng, Z. 2001. Internet GIS for public participation. Environment and Planning B: Planning and Design 28: 889-905.

Perkins, D. D., B. B. Brown, and R. B. Taylor. 1996. The ecology of empowerment: Participation in community organizations. Journal of Social Issues 52(1): 85-110.

Peterson, M. P. 2005. A decade of maps and the Internet. Proceedings of the 22nd International Cartographic Conference, July 9 to 16, 2005, A Coruña, Spain.

Pickles, J. 1997. Tool or science? GIS, technoscience, and the theoretical turn. Annals of the Association of American Geographers 87(2): 363-72.

Rambaldi, G. 2006. Gestion participative de l'information géographique et de la communication dans les pays en developpement. Electronic Journal on Information Systems in Developing Countries 25(1): 1-11.

Ramsey, K., and M. W. Wilson. 2006. Rethinking the "informed" participant: Precautions and recommendations for the design of online deliberation. In Davies, T., and B. Noveck. Eds., Online deliberation: Design, research, and practice. Stanford: CSLI.

Rinner, C., and M. Bird. 2009. Evaluating community engagement through argumentation maps—a public participation GIS case study. Environment and Planning B: Planning and Design 36(4): 588-601.

Rinner, C., C. Ke ler, and S. Andrulis. 2008. The use of Web 2.0 concepts to support deliberation in spatial decision making. Computers, Environment and Urban Systems 32(5): 386-95.

Rinner, C., J. Kumari, and S. Mavedati. 2011. A geospatial Web application to map observations and opinions in environmental planning. In Li, S., S. Dragicevic, and B. Veenendaal, Eds., Advances in WebGIS, mapping services, and applications. London: Taylor and Francis, pp. 277-91.

Robinson, P., and C. Gore,. 2005. Barriers to Canadian municipal response to climate change. Canadian Journal of Urban Research 14 (Summer 2005 Supplement): 102-20.

Rosenstone, S. J., and J. M. Hansen. 1993. Mobilization, participation, and democracy in America. New York: MacMillan.

Rowe, G., and L. J. Frewer. 2000. Public participation methods: A framework for evaluation. Science, Technology, and Human Values 25(1): 3-29.

Sani, A., and C. Rinner, 2011. A scalable GeoWeb tool for argumentation mapping. Geomatica 65(2): 145-56.

Scharl, A., and K. Tochtermann, Eds. 2007. The geospatial Web: How geobrowsers, social software and the Web 2.0 are shaping the network society. London: Springer.

Schlossburg, M., and E. Shuford. 2005. Delineating "public" and "participation" in PPGIS. URISA Journal 16(2): 15-26.

Sieber, R. 2006. Public participation geographic information systems: A literature review and framework. Annals of the Association of American Geographers 96(3) 491-507.

Sieber, R. 2008. Participatory Geoweb: A research agenda. Proceedings of the Spatial Knowledge and Information Canada Conference (SKI-Canada), Volume 2, February 14 to 17, 2008, Fernie, BC, Canada.

Taylor, D. R. F., and S. Claquard. 2006. Cybercartography: Maps and mapping in the information era. Cartographica 41(1): 1-5.

Tulloch, D. 2007. Many, many maps: Empowerment and online participatory mapping. First Monday 12(2).

Wang, W., and S. French. 2008. A multidimensional framework for facilitating wide participation and common understanding. Paper presented at the WebScience '08, Pittsburgh.

Ward, S., R. Gibson, and W. Lusoli. 2003. Online participation and mobilization in Britain: Hype, hope, and reality. Parliamentary Affairs 56(4): 652-68.

Wiedemann, P. M., and S. Femers. 1993. Public participation in waste management decision making: Analysis and management of conflicts. Journal of Hazardous Materials 33(3): 355-68.

Williams, G. 2004. Evaluating participatory development: Tyranny, power, and (re)politicisation. Third World Quarterly 25(3): 557-78.

A Place-based Tool for Assessing Cumulative Impervious Surface Outcomes of Proposed Development Scenarios

Kevin Ramsey and Aaron Poresky

Abstract: Impervious surface cover is commonly used as an environmental indicator for land-use and watershed planning. Tools for predicting the net increase in impervious surface area that will result from future land-use or development scenarios can aid planners in assessing the relative impacts on water quality, flood-control infrastructure, stream erosion, groundwater recharge, and habitat. While methodologies currently exist for estimating impervious surface cover, each has drawbacks when applied to support prospective analysis. Existing methodologies also are limited in their ability to differentiate the impacts of infill and greenfield development. To attempt to address these limitations and to fill unmet needs, a remote-sensing and regression analysis was undertaken using sample data from the U.S. National Land Cover Database (2006). The result is an impervious surface growth model capable of predicting the net increase in impervious cover at the census block group scale as a function of quantities of residential and commercial development added and relative centrality of the block group within a metropolitan regional context.

INTRODUCTION

The environmental impacts of storm-water runoff from impervious surfaces are well documented in the research literature (EPA 1992, EPA 1998, Schueler 1994, Brabec et al. 2002). Heightened concerns about these impacts have led scholars to propose impervious surface cover as a key environmental indicator for land-use and watershed planning (Arnold and Gibbons 1996). Planners seeking to take into consideration the likely storm-water runoff impacts of alternative development or land-use scenarios require tools for assessing cumulative impervious surface outcomes.

The ability to account for off-site impacts of proposed land uses is particularly important when comparing alternative development or land-use scenarios. For instance, when considered at the parcel scale, lower-density residential development often creates less impervious surface per acre than a higher-density alternative. However, compact neighborhood development creates less new impervious surface area per home because it significantly requires less miles of roadway per home (Brabec et al. 2002; EPA 2003). Better yet, infill development[1]—or building new homes within previously developed areas—takes advantage of existing services and infrastructure, further reducing the net increase in impervious area per home. Any tool that fails to account for these cumulative impacts of alternative development scenarios is at risk of providing an incomplete and misleading representation of the full environmental implications of land-use decisions.

The ability to compare the likely outcomes of alternative land-use scenarios is a topic of growing interest among planning practitioners. During the past few decades, an increasing number of metropolitan regions have engaged in visioning exercises that allocate forecasted growth to specific locations (Bartholomew 2005).[2] These projects often involve comparing a preferred vision scenario to "business as usual" in which development patterns conform to prevailing trends. Proponents of "smart growth"[3] and compact infill development also are seeking new ways to measure the environmental benefits of specific development proposals. One way to do this is to compare the impacts of the proposed infill development to the impacts of an equivalent amount of development (on a per-unit basis) occurring in conformance with prevailing development patterns (EPA 2001).[4]

There are several methods for estimating impervious surface cover based on land-use class or gross density of activity. This paper explains why each has important limitations with regard to estimating the net increase in impervious cover associated with new development. We then report on the development of a new model, dataset, and spreadsheet tool for use in assessing impervious surface impacts of proposed growth and development scenarios. This tool is designed to be practical for routine use by local planners as well as sensitive to differences in cumulative and off-site impacts associated with the location of a proposed development. We discuss tool applications as well as potential

---

[1] The term infill refers to new development that occurs within areas that are already urbanized. Infill may come in the form of building on an empty lot in an existing neighborhood or redeveloping an underutilized parcel.

[2] For a list of examples, see Urban Land Institutes "Reality Check" program, http://www.uli.org/programs/local-programs/reality-check-regional-planning-sustainable-development/.

[3] The term smart growth refers to community development and conservation strategies that promote vibrant, compact, and walkable neighborhoods while preserving natural lands and critical environmental areas, protecting water and air quality, and reusing already-developed land. See www.epa.gov/smartgrowth/about_sg.htm for more information.

[4] An example of this kind of analysis was conducted to support the Atlantic Steel Site Redevelopment Project. Read more at http://www.epa.gov/smartgrowth/topics/atlantic_steel.htm.

enhancements that could facilitate ease of integration with existing scenario planning and GIS tools.

EVALUATION OF EXISTING METHODOLOGIES

Several methodologies exist for estimating impervious surface cover based on analysis of remote-sensing imagery (Slonecker et al. 2001, Dougherty et al. 2004, Chabaeva et al. 2009). Fewer options, however, are available for predicting the impervious surface outcomes of proposed land-use or development scenarios. These options fall into two main categories. The first applies predefined impervious surface coefficients associated with individual land-use types. The second estimates impervious surface cover for census-defined areas based on activity density. Here we review this previous work and assess its suitability for supporting local and regional planning initiatives.

The most common approach used to assess the impervious surface impacts of proposed land-use scenarios is applying standardized impervious surface coefficients for designated land-use types (Brabec et al. 2002).[5] For instance, a detailed analysis of current land cover in three California metropolitan regions determined that retail land uses result in an average of 86 percent impervious land cover (Washburn et al. 2010). Using this information, California land-use planners could assume that areas zoned for retail will have approximately 86 percent impervious land cover after full build-out. This approach provides a straightforward methodology for roughly assessing future impervious surface cover based on full implementation (or build-out) of a master plan or land-use scenario. A number of planning analysis tools have adopted established coefficients to estimate the impervious surface outcomes of land-use scenarios. Examples include CommunityViz (Placeways 2013) and the Long Term Hydrologic Impact Analysis (L-THIA) spreadsheet model (Purdue University 2013).

A key limitation of this approach is that it can only be used to evaluate the outcomes of a fully implemented land-use plan. This makes it difficult to compare alternative land-use scenarios that include the same quantity of development. For instance, regional planning studies typically seek to evaluate the outcomes of growth that are expected to occur during a defined period of time (usually 20 to 40 years). To do this, planners typically allocate units of forecasted growth based on where it is anticipated or desired to occur. Local zoning or land-use plans can be used to limit the quantity of growth allocated to a specific area. But build-out is not a foregone conclusion. Therefore, land-use coefficients often prove to be impractical tools for translating forecasted growth into impervious surface outcomes.

To be comparable, alternative land-use scenarios must accommodate an equivalent amount of population, housing, and/or employment growth. This requires an impervious surface model that takes units of development as inputs (rather than land-use classes). The second category of impervious surface models fits

this description. These models estimate impervious surface cover as a function of gross population, housing, and/or employment density. Such models have been developed to estimate impervious surface cover at the scale of municipality (Stankowski 1972, Reilly et al. 2004) and census tract (Chabaeva et al. 2004).

Activity density models address the limitations of the land-use coefficient approach and therefore are more appropriate for scenario-planning studies. However, models developed to date also present some important limitations. First, municipalities and even census tracts are a fairly coarse unit of geography. Operating at as fine a geographic scale as possible is essential in scenario-planning exercises that seek to differentiate the impacts of new greenfield development at the periphery of an urbanized area from infill development. This is partly because the development density estimates become less accurate as the scale of analysis grows. Secondly, models developed to date are insensitive to location. A high-density housing development at the periphery of a metropolitan region would be expected to result in a greater net impervious surface impact than one of equal density closer to the region's core. This is because peripheral development often requires new or expanded roadways and larger parking areas due to the much higher likelihood that residents require a vehicle for daily transportation. Development closer to the core, on the other hand, would be expected to take greater advantage of existing roadways and infrastructure.

MODEL REQUIREMENTS

This study set out to develop a model, user interface, and dataset that can be used to roughly assess the net impervious surface impacts of proposed development projects. More specifically, we wanted to be able to assess the cumulative additional impervious surface cover (both on-site and off-site) that could be expected to result from a proposed development, based on the development location. Furthermore, we sought to create a tool that is both practical for routine use and can be applied anywhere in the contiguous United States. The model requirements are described in greater detail in the following section.

1. **Relevant for application throughout the United States**
   The majority of models that estimate impervious surface cover focus exclusively on a single region or state. For this study, we sought to create a model based on nationally available data that can be applied in any location within the contiguous United States. Creating a single model with nationwide scope makes it possible to execute national studies of development scenario impacts. We also sought to create a model that could be adopted for use in localities that lack the resources to create customized models based on local data and conditions.

2. **Assesses net impervious surface impacts per unit of new development**
   Assessing impervious surface impacts per unit of new development facilitates the ability to compare the relative impacts of alternative development scenarios. This interest

---

[5]   For examples, see USDA 1986, Washburn et al. 2010, SCAG 2009.

in scenario comparison grew out of work at the EPA to better understand the indirect environmental benefits of brownfield cleanup and reuse (EPA 2001, EPA 2011). This work begins with the assumption that aggregate population and job-growth projections for a given metropolitan region are independent of particular land-use policies and decisions. From this perspective, redeveloping a brownfield can be thought to displace an equivalent amount of development (in terms of housing units, commercial floor space, etc.) elsewhere in the same metropolitan region. Based on this assumption, the indirect environmental benefits (or impacts) of brownfield redevelopment can be assessed in part by comparing anticipated impervious surface growth associated with redeveloping the brownfield location to the anticipating impervious surface growth associated with an equivalent amount of development located in the fastest-growing part of the metropolitan region.[6] Crucial to this kind of analysis is the ability to measure incremental growth in impervious surface area (growth beyond current conditions). Modeling net impervious surface impact per unit of new development facilitates this kind of study.

3. **Assesses impervious surface cover as a function of development density and regional centrality**
As noted above, previous studies have shown that density of population, housing, and/or jobs can serve as reasonable predictors of existing impervious surface cover.[7] This study will take a similar approach, but with two important refinements. First, it will seek to develop a model calibrated at the smallest geographic unit possible that is supported by nationally available data—the census block group (block group). Block groups are contained within census tracts and generally contain between 600 and 3,000 people, with an optimum size of 1,500 people. Secondly, this study is interested in where proposed development sites are located within a metropolitan region. As noted in the introduction, a development site located near the center of a metropolitan region may require less new impervious surface than one at the periphery of the metropolitan region in part because peripheral locations often necessitate more driving. This is because peripheral locations often lack transportation choices and require further travel distances to reach everyday destinations. More driving means more need for pavement (per unit of development) both on-site and off-site. Therefore, this study tested additional variables representing regional centrality as well as the overall size (in terms of population and jobs) of the surrounding metropolitan region.

4. **Accounts for off-site impervious surface growth**
For reasons already stated, the ability to at least partially

account for off-site impervious surface growth is an essential feature of this model. Structuring the model to assess impacts per unit of development within a geographic area (e.g., census tract or census block group) provides nearby off-site impacts of development. (The implications of selecting a census block group level model on its ability to capture off-site impervious surface growth are discussed later in the report.)

5. **Practical for routine use**
We sought to develop a model and dataset that is ready for use in regions across the United States, without the need for additional baseline data or calibration from the local area of analysis. We also sought to develop a tool that requires only the site location and units of development as inputs, rather than fully formed land-use scenarios.

DATA SOURCES AND MODEL SELECTION

The Impervious Surface Growth Model (ISGM) that was developed from this study is a regression-based model developed to meet the needs introduced above. The selection of the form of model was based primarily on the datasets that are available to support this study and their reliability for this application. A preliminary analysis of available datasets was conducted, including correlation analyses and inspection (analytical and visual) of datasets related to their completeness and reliability. Key sources of data that were considered for this analysis are identified in Table 1, including a brief summary of their eventual use in development of the ISGM.

These datasets were reviewed and preliminary data analyses were conducted to guide interim decision making related to development of the ISGM. A summary[8] of the key practical findings of these analyses are as follows:

- ***Is it more reliable for the ISGM regression model to be based on estimates of change in input parameters over a given period (i.e., change in imperviousness from 2001 to 2006) or based on static estimates of these parameters at a "snapshot" (i.e., total imperviousness in 2006)?*** A regression based on change metrics would more directly support the estimation of net impervious surface growth (net ISG)—the net of amount of impervious surface added per incremental unit of development. However, based on findings of preliminary data analyses, a model based on static estimates was considered more reliable for the ISGM. This preference was primarily based on the observation that static estimates appear to have lower levels of relative error and "noise" than change estimates do, which, in available datasets, are based on a relatively short period of change. The result of this decision is that the model may be more reliable for estimating impervious growth in areas that have already undergone some development, as discussed further in

<hr>

[6]   A more detailed discussion of this methodological approach to assessing the impacts of brownfield redevelopment is available in EPA 2001.

[7]   In addition to Chabaeva et al. 2004, Washburn et al. 2010 estimate percent impervious surface cover at the submunicipality level based on residential density.

[8]   A more detailed account of the preliminary data analysis is available in Geosyntec 2011, https://edg.epa.gov/data/public/op/ISGM/ISGM_finalreport.pdf.

Table 1. Datasets identified and evaluated for use

Category	Description	Source(s)	Use in ISGM Development
Nationwide Land Cover and Impervious Cover Datasets	Raster datasets (30-m resolution) containing estimates of composite imperviousness, land cover, impervious cover, and changes in land cover and imperviousness from 2001 to 2006	National Land Cover Databases 2001, 2006 Multi-Resolution Land Characteristics Consortium (MRLC) (USGS, NOAA, and EPA)	Used to calculate impervious cover in each block group via spatial analysis methods in GIS
State and Local Impervious Cover Datasets	Raster datasets with various resolution and spatial extent; available from various state and local agencies	State of Massachusetts, State of Maine, State of Hawaii, State of Delaware, City of Atlanta, Oregon Metro, City of Durham, NC, King County, WA, Santa Barbara County, CA	Oregon and Massachusetts data were used to evaluate reliability of NLCD impervious cover datasets; datasets were not used directly in the ISGM
High-resolution Aerial Photography	Various datasets available at nationwide or custom extents, some with multispectral bands, circa 2009 to 2011; some sources with historical data as well	Various agencies and companies	Used to evaluate reliability of NLCD impervious cover datasets; not used directly in the ISGM
Census Block Groups	Geographic dataset of block group boundaries defined by political boundaries and population; shapefile format for geospatial analysis with other layers or tabular relationships with related datasets	U.S. Census Bureau (obtained directly from the Smart Location Database, Ramsey et al. 2012)	Used as base unit of geography for spatial analyses and regression analyses
EPA Smart Location Database[9]	A nationwide collection of population, housing, employment, transportation, and other metrics at the block group scale (e.g., population density), as well as modeled data and indices (e.g., gravity model of destination accessibility); the 2012 version of this dataset provided density metrics based on only the "unprotected" areas of each block group. See "protected areas datasets" below	EPA (Ramsey et al. 2012)	Various metrics and model results/indices from the SLI were considered in preliminary regression analyses. Certain metrics and model results from the SLI were used in the final regression model and associated user interface (See Table 2.)
Local Employment Dynamics Dataset (LED) and Longitudinal Employer-Household Dynamics Dataset (LEHD)	Information about locations of employment and residence at the census block scale (one degree finer than block group)	U.S. Census Bureau (accessed May, 2011)	Used as a source of data for estimating the number of employees in each block group
Protected Areas Datasets	Various datasets identifying locations that are restricted from development in some way, either via land cover (i.e., water) or local, state, or national planning designations (i.e., parks, national forest, military reserves, etc.)	Protected Areas Database—US (PADUS) v1.2 (USGS) (2011) Navteq land-use and water features (Navteq, 2011) NLCD 2006 (MRLC, 2011)—water land cover	Used in spatial analysis to categorize "unprotected" and "protected" portions of block groups, incorporated into 2012 SLI Database

[9]   Note that EPA's Smart Location Database was updated in 2013, after this study was complete. The 2012 release of the database is still available online. See Ramsey et al. 2012 or download the entire dataset (324 MB) at https://edg.epa.gov/data/public/op/ Smart_Location_DB_v02b.zip. For information about the latest release of the Smart Location Database, see http://epa.gov/smartgrowth/smartlocationdatabase.htm.

the "Model Validation and Reliability" section of this paper.

- ***What scale and resolution of remote-sensing analysis best balances data quality and data quantity to yield the most reliable model?*** Options considered for model development range from focused, high-resolution analysis of a relatively small number of samples (100 to 200) to a much broader analysis, considering the majority of block groups (approximately 200,000), but with estimates generated for each block group at lower resolution. Based on observations of data quality and reasonableness (above), a broad analysis was strongly preferred compared to a more focused analysis: (1) a broad range of potential independent variables (e.g., development density, destination accessibility) are likely to be needed to adequately describe the urban context, (2) regional variability may need to be considered in this or future analyses and can be much more rigorously supported by analyzing a large number of samples, and (3) observations of data quality and reasonableness indicate that the datasets that would be used in the broader analysis appear to have adequate quality and reliability.
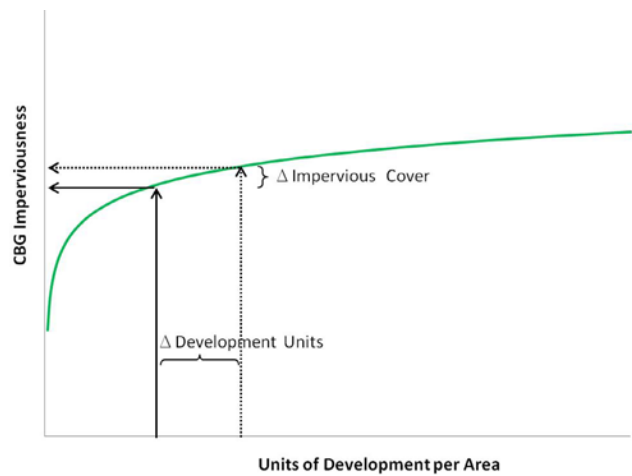
Figure 1. Conceptual model for estimation of impervious cover change

Table 2. Parameters used for regression analysis

Parameter ID	Description	Units	Source
HU2006	Housing units, estimated (2006)	Count	U.S. Census Bureau
EMP2006	Total employees; nonfederal (2006)	Count	LEHD (downloaded February, 2011)
UNP_IMP_06	Percent impervious cover in unprotected areas	%	Analysis of unprotected areas[10] and NLCD 2006 impervious cover dataset
UNP_IMPAC_06	Impervious acres in unprotected area (2006)	ac	Analysis of unprotected areas and NLCD 2006 impervious cover dataset
HU_DENS	Unprotected area housing unit density	HU/ac	Calculated from metrics above (housing units divided by unprotected area, acres)
EMP_DENS	Unprotected area employment density	EMP/ac	Calculated from metrics above (employees divided by unprotected area, acres)
D5AR*	Jobs within 30 miles, gravity weighted[11] (destination accessibility) Note: Various other parameters were evaluated as part of potential regression models that were not selected. More information about other parameters can be found in the ISGM project report (Geosyntec 2011).	jobs	U.S. Environmental Protection Agency, (Ramsey et al. 2012)

*Modeled input from Smart Location Database

[10] This analysis was concerned only with impervious land cover within the developable portion of each block group. Therefore, all areas known to be protected from residential and commercial development activity were eliminated from block group boundaries before land-cover analysis. Two national data sources were used to identify land area protected from development. NAVTEQ was used to identify city, regional, state, and national park lands. Protected Areas Dataset–U.S. (PADUS) version 1.2 was used to identify all public lands as well as private conservation lands permanently protected from development. This analysis is documented by Ramsey et al. 2012.

[11] This is a measure of "destination accessibility" and regional centrality included in the U.S. EPA's Smart Location Database (Ramsey et al. 2012). It istmeasured as the cumulative number of jobs that can be accessed from the origin census block group within a 30-mile radius, gravity weighted. Note that this metric was based on 2009 employment counts.

MODEL DEVELOPMENT

Model development consisted of (1) selecting the form of the ISGM, (2) selecting regression parameters, (3) conducting the regression analysis, (4) selecting the best-performing regression model, and (5) evaluating model reliability. The following sections describe this process.

Form of Impervious Surface Growth Model

The ISGM is based on a multivariate, nonlinear regression equation that yields an estimate of average imperviousness based on the housing unit density, employment density, and destination accessibility of the unprotected areas of each block group. This estimate of imperviousness can be multiplied by the unprotected acreage of the block group to yield an estimate of the acreage of impervious cover in the unprotected area of each block group. The hypothetical addition of development units (i.e., housing units and/or number of employees) results in adjustments to the independent parameters (i.e., increased housing unit density and/or increased employment density) in the regression, which yields an increase in the impervious cover estimated by the regression. The difference in impervious cover predicted between the baseline condition and the hypothetical adjusted condition can be attrib-

uted to the hypothetical number of units of development added. This model is conceptually illustrated in Figure 1.

Parameter Selection and ISGM Regression Analysis

The regression equation selected for use in the ISGM was chosen from a large number of potential options based on an iterative and adaptive process. Initial parameters were selected for consideration based on the results of the scatter plot matrices and nonparametric correlation analyses conducted on the preliminary dataset. Parameters were added and removed from the regression, iteratively, to attempt to improve performance. Additionally, a range of model forms was evaluated. The dataset used for the regression analysis is described in Table 2.

A stratified sampling method was used to develop and test the regression equation.

1. From a pool of all block groups in the conterminous United States that contain unprotected land area, we first excluded block groups that do not contain sufficient and consistent data on which to base the development of the regression. The resulting block group dataset used for analysis included 181,809 block groups, each containing consistent estimates of the key independent and dependent parameters.

Figure 3. Partial graphical depiction of selected regression model (D5Ar = 100,000)

Raw R-square (1-Residual/Total)	:	0.954
Mean Corrected R-square (1-Residual/Corrected)	:	0.825
R-square (Observed vs. Predicted)	:	0.827
R (Correlation coeff.)		0.909

Figure 2. Comparison of predicted to proposed imperviousness and regression statistics

2. The analysis dataset then was stratified into five equal interval bins from 0 to 100 percent impervious cover, and an equal number of random samples were selected from each bin. Stratified random sampling conducted to develop the regression model yielded approximately 25,129 samples (i.e., approximately 5,000 data points per imperviousness bin) in 37 states.

3. Using this subsample dataset, many model trials were conducted using different forms of regression equations and different combinations of potentially significant explanatory variables. The nonlinear regression modeling tool in SYSTAT© Version 12 (http://www.systat.com/) was employed to find the best combination of coefficients for each trial and generate regression statistics. These statistics were evaluated along with an inspection of scatter plots of the predicted imperviousness versus measured imperviousness (NLCD 2006) for each trial. Based on these trials, a best-performing regression equation was identified.

Best-performing Regression Equation
The best-performing nonlinear regression model that was obtained has the following form and coefficients.

$$\%IMP = \frac{100}{1 + \dfrac{1}{0.008 + 0.1227 \times HU_{UAC} + 0.093 \times EMP_{UAC} + 0.000000739 \times D5AR}}$$

Where: %IMP is percent imperviousness of the unprotected area of the block group

HU_{UAC} is the housing units per unprotected acre

EMP_{UAC} is the employees per unprotected acre

$D5AR$ is number of jobs within 30 miles based on a gravity model

Figure 2 displays the comparison of impervious cover "predicted" by the best-performing regression model to the "actual" imperviousness measured by the 2006 NLCD. Figure 3 depicts the regression equation graphically for an example "solution surface" holding the D5AR variable to 100,000 jobs.

MODEL VALIDATION AND RELIABILITY
Model validation was an integral element of developing the regression model, and was part of the iterative process used to develop the selected model. The model was validated in three primary ways, as described in the paragraphs below.

Application to Remaining Sample Data
The selected regression model was applied to the remaining 156,520 samples (block groups) that were not used in the development of the model. This validation was based on a comparison made between the residuals of the model development dataset (25,129 block groups, Figure 4) and the residuals of the remaining dataset (156,520 block groups, Figure 5). Residuals are fairly evenly distributed for both datasets, and the mean and median of residuals differ by only 1 percent to 2 percent imperviousness between the datasets–the standard deviations differ by less than 1 percent. These differences can likely be attributed to the greater influence of the middle of the range of imperviousness (30 to 60 percent) in the full dataset compared to the stratified model development subsample, as well as the presence of potential outliers. A truly normal distribution will have a skewness of zero and kurtosis of three. As shown in Figure 4 the skewness is only slightly negative and the kurtosis is slightly higher than three. While normally distributed residuals are preferred in regression analysis, residuals that are approximately normal and have approximately constant variance indicate that the regression equation will produce reasonably accurate predictions (Helsel and Hirsch 2002). This comparison indicates the model development subsample is reasonably representative of the full population.

	RESIDUAL
N. of Cases	25,129
Minimum	-95.753
Maximum	72.777
Median	-0.821
Arithmetic Mean	-0.543
Standard Deviation	11.967
Skewness (G1)	-0.061
Kurtosis (G2)	3.080
Anderson-Darling Statistic	201.184
Adjusted Anderson-Darling Statistic	201.190
p-value	<0.01
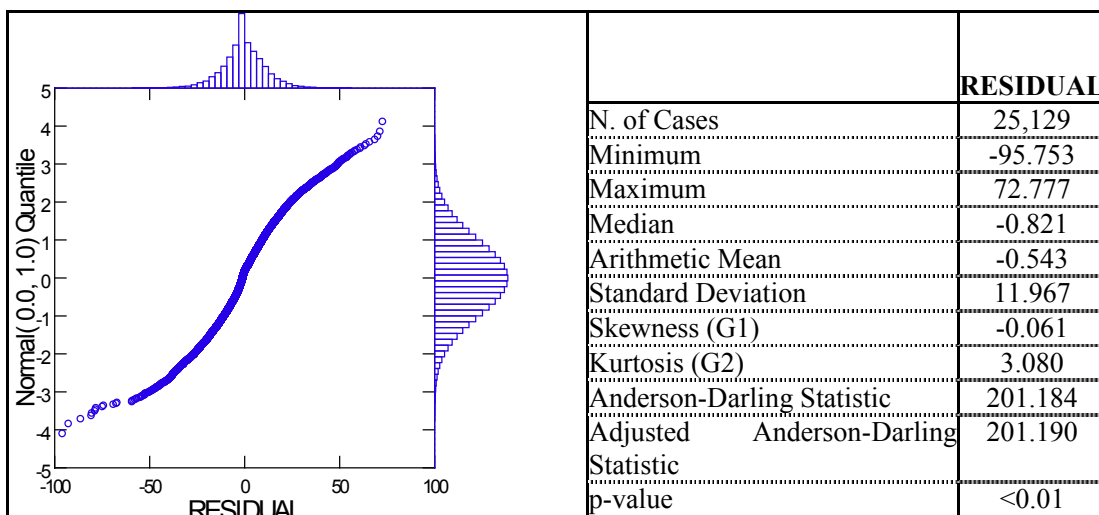
Figure 4. Residual statistics for data used in regression model

	RESIDUAL
N of Cases	156,520
Minimum	-98.817
Maximum	75.949
Median	-1.920
Arithmetic Mean	-3.144
Standard Deviation	10.896
Skewness (G1)	-0.700
Kurtosis (G2)	5.094
Anderson-Darling Statistic	29,289
AdjusteAnderson-Darling Statistic	29,289
p-value	<0.01
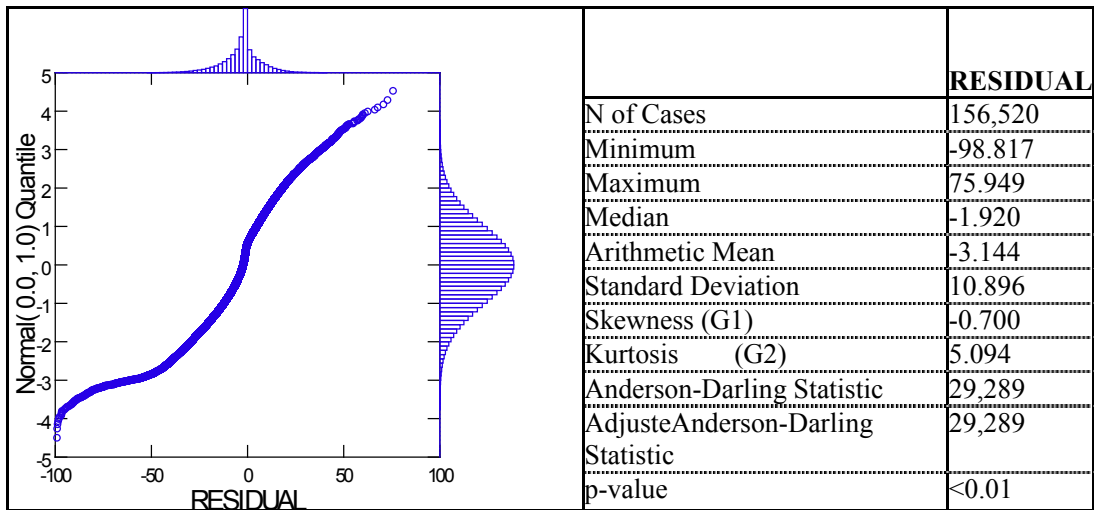
Figure 5. Residual statistics for remaining data not used in regression model

Comparison to Similar Independent Study

The relative error, variability, and magnitude of predictions from the best-performing regression equation were compared to a recent comparable effort by the state of California (Washburn et al. 2010). The California analysis used high-resolution remote sensing of randomly selected neighborhoods in several cities to estimate the imperviousness of a range of land uses in California. The sample set included more than 330 residential neighborhoods at densities ranging from 1 to 50 dwelling units per acre as well as a variety of other neighborhoods that were not classified by an analogous density metric. Among other outcomes, the analysis yielded a regression equation that can be used to correlate land-use imperviousness to housing unit density for residential land uses. Figure 6 shows the plot of imperviousness versus housing unit density derived from this analysis. For comparison, the ISGM regression model is overlaid on this chart (holding employment at 0 and D5Ar at the approximate median value of 100,000).

While these regressions are not directly comparable (block groups are generally at a larger scale and less homogenous than the neighborhoods surveyed), the relative magnitudes and shapes are similar. The ISGM equation appears to fit the California data fairly well, and the regression statistics of the ISGM equation (based on fit to nationwide block groups) compares favorably to the best fit that was found for the California ISC analysis (based on California neighborhoods).

Reasonableness Inspection of ISGM Predictions

The ISGM was applied to a subset of block groups to predict the net ISG associated with hypothetical increases in housing units and employees. Twenty-four block groups from five U.S. cities were studied. These block groups were selected prior to application of the model to represent a cross section of block groups from different locations within the urban context (i.e., downtown versus suburban), different city sizes, and states with different land-use management policies. Net impervious surface
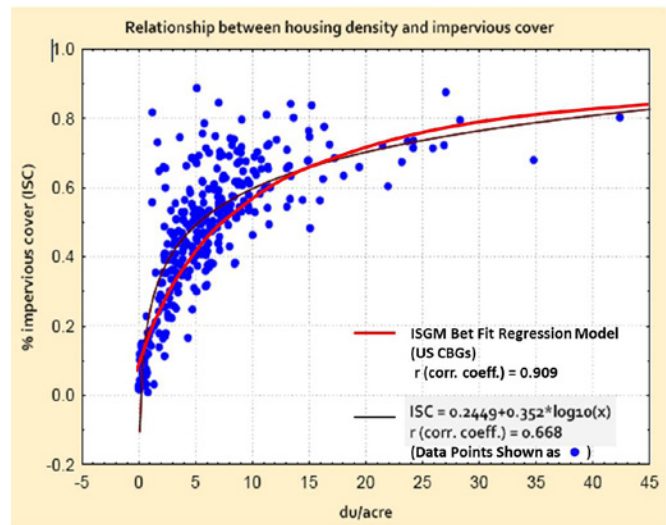
Figure 6. Comparison of ISGM results to California ISC analysis
Note: The correlation coefficient for the ISGM best fit regression model is based on its fit to the selected subsample of nationwide block groups for comparison; it is not based on the California land-use data that is plotted on this chart.

growth per additional unit of development was estimated based on a nominal increase in development units of 100 units. Figure 7 shows an example case study block group from this reasonableness evaluation.

This inspection of multiple case study applications showed that results are reasonable and followed expected trends. Of the block groups inspected, the net residential ISG ranged from approximately 4,000 square feet per housing unit in urban fringe block groups to approximately 200 square feet per housing unit in highly urbanized block groups. Net employment ISG followed a similar trend to net residential ISG with somewhat lower values predicted. This is expected based on the form of the regression equation and appears to yield reasonable results in the block
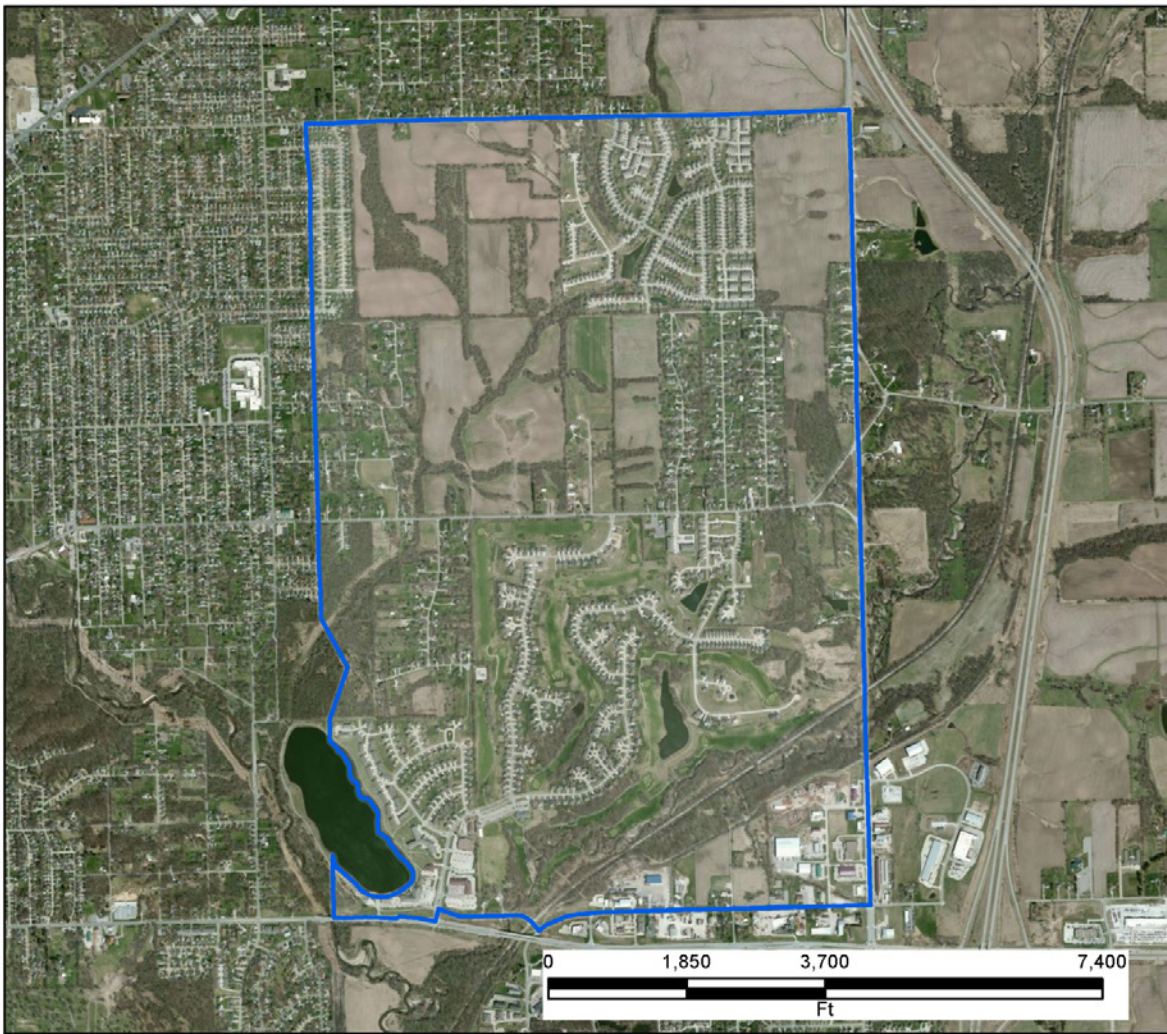
CBG ID: 191530106007

CBG: 191530106007
Des Moines-West Des Moines, IA
Baseline hu/ac: 0.7
Baseline emp/ac: 0.38
D5Ar: 123787
Net ISGr = 0.082 IAC/hu
(3568 ISF/hu)
Net ISGe = 0.062 IAC/emp
(2708 ISF/emp
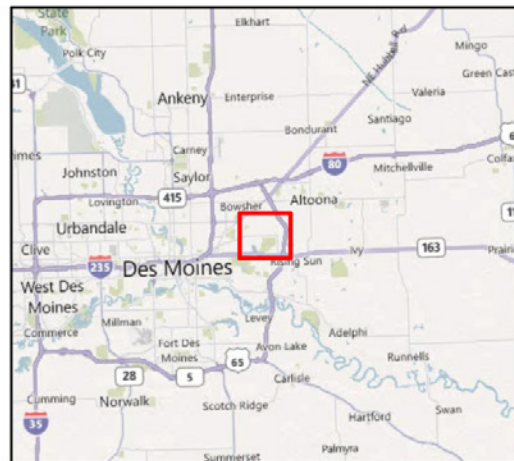
Figure 7. Case study application of ISGM to an example block group

Table 3. ISGM user interface fields

Field Type	Field ID	Field Description	Units	Source
User Input	CBG	Block group ID	text	User entered
	MSA	Metropolitan statistical area	text	Returned via lookup from ISGM Database based on block group ID Primary Key
	ADD_HU	Added housing units	hu	User entered
	ADD_EMP	Added employment units	jobs	User entered[12]
	ADD_Protected	Added acres of land protected from development	acres	User entered
Block Group Baseline Conditions	UNP_ACRES	Best estimate of unprotected area, ac	acres	Returned via lookup from ISGM Database based on block group ID Primary Key
	HU_DENS	Housing unit density (unprotected, baseline, 2010)	hu/acre	Returned via lookup from ISGM Database based on block group ID Primary Key
	EMP_DENS	Employment density (unprotected, baseline, 2009)	jobs/acre	Returned via lookup from ISGM Database based on block group ID Primary Key
	D5AR	Jobs within 30 miles, gravity weighted (2009)	jobs	Returned via lookup from ISGM Database based on block group ID Primary Key
Development-adjusted Block Group Conditions	HU_DENS_ADJ	Housing unit density (unprotected, adjusted)	hu/acre	Calculated based on 2010 conditions plus user-entered number of added housing units and added protected area
	EMP_DENS_ADJ	Employment density (unprotected, adjusted)	jobs/acre	Calculated based on 2009 conditions plus user-entered number of added jobs and added protected area
	D5AR_ADJ	Jobs within 30 miles, gravity weighted (D5Ar, adjusted)	jobs	Calculated based on 2009 D5ar plus user-entered number of added jobs
Results	ISG_NET	Net impervious surface growth	acres	{ISGM IMP (Adjusted) - ISGM IMP (Baseline)} See note[13]
	ISG_MAX	Maximum possible impervious surface growth in 2006	acres	Remaining impervious surface in block group (NCLD 2006). Value displayed if ISG_NET > ISG_MAX
	QUAL	Qualifier	text	Returns qualifying information where model predictions as applicable
	NOTES	Notes about results	text	Returns notes, as applicable.

[12] This field does not refer to jobs associated with construction. Rather it refers to the total number of additional people who are estimated to be working in the block group after the new construction is complete.

[13] ISGM IMP (Baseline) = Block group unprotected area impervious area predicted for the baseline (2009-2010) condition based on the ISGM regression equation using the baseline independent input variables. ISGM IMP (Adjusted) = Block group unprotected area impervious area predicted for the development-adjusted condition based on the ISGM regression equation using the development adjusted independent input variables.

groups inspected. While the magnitudes are reasonable, specific examples were observed where the regression may not fully describe the expected variability.

Summary of Validation and Limitations

Overall, the ISGM appears to be a valid basis for estimating net impervious surface growth across a wide range of urban, suburban, and rural conditions. While the model may overpredict or underpredict imperviousness at a block group level, it appears to provide a reasonably reliable estimate of relative net ISG, on average. However, four key limitations should be understood in applying the model:

- First, the model does not account for vacancy in commercial buildings. Using employment density as a proxy for commercial activity presents an inherent limitation to the model, which is most acute in areas with a great deal of vacant office or retail space. In such locations, the model would tend to be biased toward lower estimates of static imperviousness in the baseline condition than was actually present. In these cases, the net impervious surface growth predicted by the model would tend to be overestimated.

- Second, while the model accounts for impervious surface growth associated with off-site transportation infrastructure that is collocated within the same census block group, it does not account for impervious surface growth associated with transportation infrastructure outside of the same block group. For instance, a new highway built to serve a rapidly growing suburban area would likely increase impervious surface cover in areas outside of the block groups in which the rapid development is occurring. In these situations, the total net impervious surface growth associated with new development could be underestimated at the block group level. However, this issue is mitigated in part by the facts that units of census geography generally are much larger in lower density areas at the periphery of a metropolitan region—the very places where one may anticipate off-site impervious surface growth to be the greatest. With larger units of geography, more off-site impacts will be captured.

- Third, this model underestimates impervious surface cover in smaller block groups that have a large proportion of unprotected land cover devoted to transportation infrastructure. Examples could include an urban railyard or port industrial district or an urban block group bisected by a highway. In these cases, the model would tend to be biased toward lower estimates of static imperviousness than was actually present. This has the effect of predicting greater net ISG with added development units than would actually be expected and could result in some systematic overestimation of impervious surface growth associated with new development.

- Fourth, because the model is based on static estimates of imperviousness previous rather than change estimates, the model is inherently based on trajectories of neighborhood densification that represent past development patterns across the United States.. The model thus assumes that future development will follow similar densification patterns. So, for example, the model inherently assumes that new residential development at the outer periphery of a metropolitan region will be relatively low in density (as this is the prevailing development patterns in regions across the United States). This model does not account for factors such as local zoning or urban growth boundaries that may cause new development to deviate from these prevailing patterns. However, the destination accessibility variable (D5ar) does serve as an indicator of the "centrality" of the location and implicitly accounts for some of the factors that influence decisions about the type of development that will occur.

- Finally, and related to the fourth limitation above, the model does not account for innovative new development practices intended to minimize impervious surface cover. For instance, new residential neighborhoods with smaller lot sizes, narrower street widths, and a mix of land uses that promote walkability can potentially result in less impervious surface growth, per unit, than conventional large lot residential development. However, because the model works at the block group scale, it cannot account for density of development at the scale of a subdivision or development site. In other words, it cannot differentiate between two development proposals for a single block group—unless one proposal formally sets aside acreage as protected from development (essentially allowing the analyst to adjust the density associated with the remaining area inside the block group).

IMPERVIOUS SURFACE GROWTH MODEL TOOL

We developed a simple spreadsheet-based tool to provide access to the ISGM algorithms and to facilitate evaluation of the predicted effect of proposed development on net impervious surface growth. The interface consists of a form in Excel 2007 with fixed columns and an expandable number of rows. Each row can be used to estimate the net ISG based on a user-defined block group and a user-defined increase in units of development. Table 3 describes the fields in the tool and the algorithms used to return the estimated value. Full documentation of methods, limitations, and user instructions are provided in the Technical Report describing the development of the ISGM (Geosyntec 2011).

The ISGM User Interface is intended to allow bulk entry of block group development scenarios and return estimates of the net ISG associated with each scenario. For each row, the spreadsheet returns the estimated net impervious surface growth. The current version can support simultaneous computation of results of up to 25,000 scenarios.

CONCLUSION

In conclusion, we discuss how effectively the ISGM addresses requirements laid out at the beginning of this paper. We also discuss the potential suitability of the ISGM for various applications

in watershed and land-use planning. Finally, we offer thoughts regarding enhancements that could extend the functionality of the ISGM in fruitful directions.

Model Requirements

We believe the ISGM represents a significant advancement in meeting the unmet scenario analysis needs described earlier in this paper.

- *Relevant for application throughout the United States.* The ISGM supports scenario analysis throughout the contiguous United States. Hawaii and Alaska were excluded from the modeling because of land-cover data availability.
- *Assesses net impervious surface impacts per unit of new development.* The ISGM returns an estimate of the net impervious surface growth per change in units of housing units and employees.
- *Assesses impervious surface cover as a function of development density and regional centrality.* The ISGM input parameters include development density (housing units per unprotected acre and employees per unprotected acre) and jobs within a 30-mile radius (an indicator regional centrality).
- *Accounts for off-site impervious surface growth.* The ISGM implicitly accounts for off-site impervious surface growth (e.g., roads, other infrastructures) that is within the block group where development occurs. It does not attempt to account for off-site impervious surface growth that may occur in other block groups.
- *Practical for routine use.* The ISGM interface has been developed to provide simple access to the ISGM and allow a large number of scenarios to be processed efficiently.

Model Reliability and Intended Uses

Although limitations have been identified, the ISGM generally is considered to provide reliable estimates of net impervious surface growth to support planning-level scenario analysis across a wide range of urban, suburban, and rural conditions. The model may overpredict or underpredict imperviousness at a block group level.

Potential Extended Applications

Give the importance of impervious cover and impervious cover growth in water resources applications, the tool is expected to have applications beyond its original intended functions.

- *Development site-selection analysis.* While more detailed site-specific analysis would always be required to fully understand the impacts of a proposed development project, the ISGM has the potential to allow users to quickly and roughly compare the estimated impervious surface impacts of a number of proposed development sites. Users of such information might include developers, urban planners evaluating development proposals, or citizens concerned about the impacts of proposed development on water quality.
- *Growth planning and impact analysis.* The ISGM has the potential to allow urban planners and policy makers to

conduct rapid planning level analysis of the relative water quality impacts of various development and land-use scenarios. Given a regional growth projection in terms of numbers of new housing units and numbers of new jobs, the ISGM could be used to rapidly evaluate the comparative impacts of various growth management scenarios on impervious surface growth and (with further analysis) water quality. This information could be used in conjunction with information from other tools (e.g., estimates of vehicle miles traveled) to identify growth scenarios that minimize impacts.

- *Watershed and drainage planning.* Based on land-use policies and population growth estimates, the tool could be used to generate long-range estimates of impervious surface growth at a watershed or subwatershed scale. This information could be used to help identify receiving waters that are most likely to be impacted by future development, which, in turn, could be used to prioritize monitoring activities to collect baseline data. This information also could be used in drainage master planning to identify long-range needs for improvements to major drainage infrastructure to support future development.
- *Other potential uses.* Given the importance of impervious cover in storm-water planning, a variety of other potential uses may exist for the ISGM or the underlying regression model. For example, the regression model developed as part of the tool has potential to be used to improve estimates of impervious cover of various types of development.

Potential Enhancements

A number of potential enhancements currently are under consideration to improve the ISGM.

- *Translating output into percent impervious cover.* A simple extension of the ISGM interface could enable output to be translated output in terms of percent impervious cover. This currently is supported via postprocessing methods.
- *Integration into established GIS-based scenario planning tools.* The ISGM could be readily incorporated into other tools used for scenario planning, such as the EPA BASINS (Better Assessment Science Integrating point and Nonpoint Sources) program.
- *Ability to calculate impervious cover by watershed for land-use scenarios.* The ISGM currently provides estimates by block group. However, watershed boundaries do not necessarily align with block group boundaries. Incorporating a GIS interface for the ISGM could enable estimates to be generated for watershed boundaries.
- *Improvements utilizing impervious cover change datasets.* When impervious cover change estimates become available over a longer time window (i.e., release of newer versions of the NLCD for comparison with 2001 NLCD), the relative error in these estimates may be smaller relative to the magnitude of changes that have occurred over this longer time period. With improved reliability in change datasets, it may be possible to enhance the ISGM, particularly in suburban/urban fringe areas.

Acknowledgments

About the Authors

Kevin Ramsey is a Policy Research Fellow in the U.S. Environmental Protection Agency's Office of Sustainable Communities. He oversees the development of GIS data products and tools that enable performance evaluation of alternative land-use and development scenarios. He also serves on the HUD-DOT-EPA Partnership for Sustainable Communities performance measurement work group. He received his Ph.D. in geography from the University of Washington.

U.S. Environmental Protection Agency
Office of Sustainable Communities
Phone: (202) 566-1153
E-mail: Ramsey.Kevin@epa.gov

Aaron Poresky is a water resources engineer with Geosyntec Consultants in Portland, Oregon, where he focuses on water resources impact analysis and planning, storm-water facility design, and applied research. He holds degrees in civil engineering (B.S.) and environmental engineering (B.S.) from Oregon State University.

Geosyntec Consultant
Phone: (971) 271-5891
E-mail: APoresky@Geosyntec.com

References

Arnold, C. L., and C. J Gibbons. 1996. Impervious surface coverage: The emergence of a key environmental indicator. Journal of the American Planning Association 62(2): 243-58. Doi: 10.1080/01944369608975688.

Bartholomew, K. 2005. Integrating land use issues into transportation planning: Scenario planning. University of Utah and FHWA, 34. Http://faculty.arch.utah.edu/ bartholomew/SP_SummaryRpt_Web.pdf.

Brabec, E., S. Schulte, and P. L. Richards. 2002. Impervious surfaces and water quality: A review of current literature and its implications for watershed planning. Journal of Planning Literature 16(4): 499-514.

Chabaeva, A. A., D. L. Civco, and S. Prisloe. 2004. Development of a population density and land use based regression model to calculate the amount of imperviousness. Paper presented at the ASPRS Annual Conference Proceedings, Denver, Colorado.

Chabaeva, A., D. L. Civco, and J. D. Hurd. 2009. Assessment of impervious surface estimation techniques. Journal of Hydrologic Engineering 14(4): 377-87.

Conway, T. M., and R. G. Lathrop. 2005. Alternative land use regulations and environmental impacts: Assessing future land use in an urbanizing watershed. Landscape and Urban Planning 71(1): 1-15. Doi: 10.1016/j.landurbplan.2003.08.005.

Dougherty, M., R. L. Dymond, S. J. Goetz, C. A. Jantz, and N. Goulet. 2004. Evaluation of impervious surface estimates in a rapidly urbanizing watershed. Photogrammetric Engineering and Remote Sensing 70(11): 1275-84.

EPA. 1992. Environmental impacts of storm water discharges—A national profile. U.S. Environmental Protection Agency.

EPA. 1998. National water quality inventory: 1996 report to Congress. U.S. Environmental Protection Agency.

EPA. 200. Comparing methodologies to assess transportation and air quality impacts of brownfields and infill development: Development, community, and environment division. U.S. Environmental Protection Agency.

EPA. 2001. Our built and natural environments. Washington, D.C.

EPA. 2009. Land-use scenarios: National-scale housing-density scenarios consistent with climate change storylines (final report). EPA/600/R-08/076F, 2009.

EPA. 2011. Air and water quality impacts of brownfields redevelopment: A study of five communities. Washington, DC: U.S. Environmental Protection Agency.

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham. 2011. Completion of the 2006 national land cover database for the conterminous United States. PE&RS 77(9): 858-64. Http://www.mrlc.gov/nlcd2006.php.

Geosyntec Consultants. 2011. Dataset for brownfield air and water quality impact evaluation: Impervious surface growth model. Https://edg.epa.gov/data/public/op/ISGM/ ISGM_finalreport.pdf.

Helsel, D. R., and R. M. Hirsch. 2002. Statistical methods in water resources techniques of water resources investigations. Book 4, Chapter A3. U.S. Geological Survey, 522 pages.

LEHD, 2011. Longitudinal employer-household dynamics. Accessed online at http://lehd.did.cen sus.gov/led/.

Placeways. 2013. CommunityViz. Accessed November 12, 2913, at http://placeways.com/ communityviz/.

Purdue University. 2013. Long-term hydrologic impact analysis (L-THIA). Accessed October 17, 2013, at https://engineering.purdue.edu/~lthia/.

Ramsey, K., D. Theobald, and J. Thomas. 2012. EPA's smart location database: A national dataset for characterizing location sustainability and urban form. Washington, DC: U.S. Environmental Protection Agency, https://edg.epa.gov/data/Public/OP/SLDv02_docs.zip.

Reilly, J., P. Maggio, and S. Karp. 2004. A model to predict impervious surface for regional and municipal land-use planning purposes. Environmental Impact Assessment Review 24(3): 363-82. Doi: 10.1016/j.eiar.2003.10.022.

SCAG. 2009. Southern California Association of Governments SB 375: Conceptual land-use scenario methodology. Southern California Association of Governments.

Schueler, T. 1994. The importance of imperviousness. Watershed Protection Techniques 1(3): 100-11.

Slonecker, E. T., D. B. Jennings, and D. Garofalo. 2001. Remote sensing of impervious surfaces: A review. Remote Sensing Reviews 20(3): 227-55.

Stankowski, S. J. 1972. Population density as an indicator of urban and suburban land-surface modification. USGS Professional Paper 800-b: B224.

Tang, Z., B. A. Engel, B. C. Pijanowski, and K. J. Lim. 2005. Forecasting land-use change and its environmental impact at a watershed scale. Journal of Environmental Management 76(1): 35-45. Doi: 10.1016/j.jenvman.2005.01.006.

Urban Land Institute. 2013. Reality check. Retrieved October 18, 2013, from http://www.uli. org/programs/local-programs/reality-check-regional-planning-sustainable-development/.

USDA. 1986. Urban hydrology for small watersheds. U.S. Department of Agriculture Technical Release TR 55. Washington DC: USDA Natural Resources Conservation Office, Conservation Engineering Division.

Washburn, B., K. Yancey, and J. Mendoza. 2010. User's guide for the California impervious surface coefficients.. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, http://oehha.ca.gov/ecotox/iscug123110.html.

The Role of Collaboration in Spatial Data Infrastructures

Watse T. Castelein, Arnold K. Bregt, and Lukasz Grus

Abstract: *This research presents an exploratory study on the factors and dynamics of collaboration processes in the context of spatial data infrastructures (SDIs). We explored the role of collaboration in the national SDIs of the Netherlands and Spain. In interviews, key stakeholders were questioned about collaboration factors and the dynamics in their national SDIs. A rather similar pattern of evolving collaboration could be identified, with more stakeholders becoming involved and a collaboration process that becomes better organized and more formalized. Our research indicates that the SDIs are getting more embedded in administrative organizational and legal structures, relying on the distributed competences and knowledge of the different stakeholders. However, the attitude and level of engagement of SDI stakeholders remains a crucial factor for collaboration. Therefore, a good balance should be found between establishing top-down formalized collaboration structures and informal, more spontaneous ways of collaboration.*

INTRODUCTION

Geoinformation has become indispensable for solving issues concerned with public safety, spatial planning, the environment, and providing e-services to citizens and companies (European Commission 2007, VROM 2008). Many initiatives since the early 1990s aimed at increasing the availability and accessibility of geographic information through the development of spatial data infrastructures (SDIs) (Onsrund 1999, Masser 1999, Crompvoets 2006). These initiatives seek to facilitate accessing and sharing of spatial data, to reduce the duplication of spatial data collection by both users and producers, and to enable better utilization of spatial data and associated services (Grus, Crompvoets, and Bregt 2010). It is believed that through well-established and properly functioning SDIs, the general economic, social, and environmental benefits can be realized (Masser 2007). SDIs have the potential to spatially enable governments by providing better service to decision makers, politicians, and societies (Rajabifard et al. 2003, Masser et al. 2008). Nonetheless, SDIs are facing challenges to attract users and to meet the requirements of their stakeholders (Georgiadou et al. 2010, Nedović-Budić et al. 2008, Budhathoki et al. 2008).

Several authors have suggested that stakeholder collaboration plays a key role within SDIs (Nedović-Budić and Pinto 2000, Warnest 2005, McDougall 2006). Within SDIs, different actors must work together, including planners and decision makers, data collectors, and analysts (De Man 2013). Involved stakeholders need to share experiences and resources to develop SDIs (Akinyemi 2011). Nevertheless, SDIs often are hampered by fragmentation and lack of collaboration between stakeholders (Thellufsen et al. 2009, De Andrade et al. 2011). SDIs remain complex because of the great variety and large number of stakeholders and their different needs (Grus et al. 2010). Moreover, the development of SDIs is a dynamic process (Koerten 2011). As SDIs emerge, the number of stakeholders involved and the relations between them increases. Organizational structures to define SDI policies and practices are changing, emphasizing partnerships, social networks, user participation, and multisectoral collaboration (Craglia and Annoni 2007, Budhathoki et al. 2008, Díaz et al. 2011). Nevertheless, little research has been conducted that looks explicitly at critical aspects for stakeholders collaboration and evolving dynamics of collaboration processes. More effort is needed to examine stakeholder interaction and collaboration processes within SDIs (Mc Dougall 2006, Elwood 2008, Vandenbroucke et al. 2009).

The term *collaboration* is ambiguous but generally is defined as stakeholders working together toward a shared goal. By working together, individual entities can pool scarce resources and duplication of services can be minimized to achieve an objective that would not otherwise be possible to obtain as separate actors working independently (Gadja 2004, Frey et al. 2006). Collaboration processes have been analyzed in different branches of science, including public management (Ansell and Gash 2008, Daley 2009, Navarrete et al. 2010), organizational science (Podolny and Page 1998, Todeva and Knoke 2005), and business management (Powell 1990, Camarinha-Matos and Afsarmanesh 2006, Allee 2008). Analyzing critical factors and evolving dynamics can help to understand and evaluate collaboration processes and to develop effective collaboration strategies (D´amour et al. 2005, Ødegard, 2006, Fletcher et al. 2009). Much depends on the purpose and the application domain—which factors are relevant and how collaborations evolve (San Martin-Rodriguez et al. 2005).

This research presents an exploratory study on the factors and dynamics of collaboration processes in the context of SDIs. We explored collaboration within two case studies: the national SDIs of the Netherlands and Spain. Our overall aim is to gain a better understanding of critical collaboration factors for the development and implementation of SDIs. In the next sections, we will further elaborate on the applied research methodology and the case studies will be described. We then will present our research

findings, reflect on them, and draw more general conclusions.

RESEARCH METHOD

To explore the factors and dynamics of SDI collaboration processes, we applied an interpretative approach aiming at theory building, based on an inductive approach (Yin 2003, Paré 2004, Andrade 2009). Interpretative research can help researchers understand human thought and action in their social and organizational context (Klein 1999). Insights are derived from the specific phenomena studied, to illuminate particular features and patterns (Neuvel 2009). It moves from specific observations to broader generalizations and theories. Case studies are an interesting research method in interpretative research, for they can provide in-depth understanding of specific phenomena (Benbasat et al. 1987, Yin 2003). In addition, through case studies, phenomena can be studied in their context and, therefore, can provide context-dependent knowledge (Paré 2004). Because our study aims at a more in-depth understanding of SDI collaboration in the social and organizational context of the SDI development, an interpretative case-study approach was considered appropriate for this research.

Data to explore SDI collaboration processes were collected in two case studies. The Netherlands and Spain were selected as case studies, because of the familiarity of the authors with the SDIs, the availability of documents, and the possibility to interview key stakeholders. Both countries have well-established national SDIs with the engagement of many stakeholders (SADL 2011a, b). On the other hand, institutional and organizational structures of both countries are different. This offers the opportunity to reflect and compare as well generic collaboration factors as context specific factors.

For data collection, semistructured interviews were conducted in April and May of 2012 for both case studies. The familiarity of the authors with both SDIs enabled them to identify, on the basis of personal contacts and available documentation, individuals with central roles in the coordination of the implementation process. For both SDIs, six experts were initially approached, of whom 11 finally participated in our semistructured interviews. Taking into consideration their positions, experience, and familiarity with the topic, they have an accurate understanding of their organizations' positions and a general overview about stakeholder collaboration within their SDI. They included the national SDI policy coordinator, an initiator of early SDI development, a key technical SDI coordinator, a representative from data providers, a regional SDI coordinator, and an SDI research coordinator. This provided us with different perspectives and enabled a more in-depth understanding of SDI collaboration.

The 11 interviewees were interviewed about: their own roles and motivation to be engaged in the SDI development; the importance, drivers, and dynamics of the SDI collaboration process in their national SDIs; and their perspectives on critical collaboration factors for the development and implementation of the national SDI. Table 1 presents the interview questions that served as guidelines. However, the order of the questions asked depended on the responses of each interviewee. The interviews took about one hour each.

Table 1. Interview items and questions

Item	Questions
Own role and motivation to be engaged in the SDI development	In which ways were you personally involved in the development of the SDI? What was your main motivation to get involved? Which organization do you represent? Why it is important for your organization to be involved in the development of the national SDI? What are the benefits for you organization?
The importance, drivers, and dynamics of the SDI collaboration process in their national SDI	What has been the role of collaboration in the SDI development? Why stakeholders thought it was good to start to collaborate? How contacts between stakeholders have been established and what were the main issues in the beginning? Was there a clear strategic plan for the development of the SDI? What was the role of more "spontaneous" bottom-up initiatives in the development of the SDI? How did collaboration evolve? Did collaboration become more structured/organized? What were the drivers for these changes?
Perspective on critical SDI collaboration factors	What is according to you critical for stakeholder collaboration in SDIs? What are the main barriers for good collaboration? Did collaboration depend on a few leaders/personal contacts? What type of dynamic do you see in the collaboration process between SDI stakeholders? How do you see the dynamic between informal contacts and more structured/organized collaboration? Was there a growing need to formalize the collaboration?

In our analysis of the results, to clarify how our conclusions were derived, citations were used to make arguments of interviewees explicit. Furthermore, the data collected through interviews were validated with documentary evidence in the form of policy documents, monitoring reports, and academic work dealing with the cases involved. As a first step toward our analysis of collaboration within SDIs, we start, in the next section, with a description

of the development trajectory and the organizational context of each case study.

CASE STUDY DESCRIPTIONS

The Spanish National SDI
In Spain, the national SDI was initiated in 2002 with the establishment of a working group for the definition and development of the "Infraestructura de Datos Espaciales de España." Since then, the SDI has been developed and implemented under the supervision of the Geographic High Council, supported by the National Geographic Institute, several ministries and regional and local governments. Central to the Spanish national SDI has been the development of a national geoportal, which has been online since July of 2004 (SADL 2011a). The national geoportal integrates servers, services, nodes, geoportals, and resources of different SDI initiatives in Spain. Originally, the Spanish SDI has been launched with no fixed regulations, but in 2007 an organizational structure for public geographic data and services providers on the national level was established with the approval of the Royal Decree 1545/2007. A second legal framework was approved in 2010, transposing the European INSPIRE directive (European Commission 2007) into a national law (Jefatura del Estado 2010). This law is obligatory for stakeholders at all administrative levels and provides the national SDI with a strong legal basis. With this legal framework, the Geographic High Council acts as a management board for the SDI, watching the implementation of the development of the national SDI, with a specific focus on INSPIRE (SADL 2011a).

The basic philosophy of the national Spanish SDI is to create an SDI where all levels of government can share their geographic information and make it available for the citizens. The idea from the beginning was that regional/local governments needed to set up their own SDIs, which were integrated to create the national SDI. On the national level, the servers, services, nodes, geoportals, and resources of distinct SDI initiatives are integrated, creating an interoperable infrastructure (Mezcua-Rodríquez 2009). Participation implies free access and reuse of network data and services for the participants. The national Spanish SDI tries to involve all the relevant stakeholders in the Spanish GI sector. National, regional, and local government, universities, and the private sector are participating in the development of the SDI (SADL 2011a). Working groups are guiding the implementation with the participation of 165 individual members from more than 60 organizations. However, not all stakeholders on the different administrative levels are participating in the development of the SDI.

The Dutch National SDI
The development of the Dutch SDI dates back to 1990 when Ravi, a network organization for geoinformation, was established. Initially, Ravi was an official advisory committee on land information for the Ministry of Housing, Spatial Planning and Environment. In 1993, it became an independent consultative body for geoinformation, its members being representatives from various public sector bodies (Van Loenen and Kok 2002). This led to the publication of the Ravi structure plan for land information, which can be seen as the initiation of the Dutch SDI (Kok van Loenen 2005). In 1995, Ravi extended this vision, which initiated the start of the National Clearinghouse Geo-Information in 1997, a metadata catalogue describing geodatasets owned by the participating SDI stakeholders (Koerten 2011). In 2007, Ravi and the national clearinghouse foundation merged to form Geonovum. Since then, Geonovum acts as the executive SDI committee in the Netherlands with the task of coordinating the development of the SDI and providing better access to geoinformation in the public sector. On a strategic level, the Geo-information Council, established in 2006, advises the Ministry on strategic actions relating to the geoinformation sector (Grus et al. 2010).

Since 2008, the Dutch SDI is being constructed by implementing the vision and strategic plan called GIDEON (VROM 2008). The document has been developed in close cooperation with 21 stakeholders and aims to develop a key geoinformation facility for the Netherlands that all parties in Dutch society will be able to use. GIDEON establishes four goals and seven implementation strategies that were intended to be realized by 2011. Various parties have been working together to execute GIDEON. The implementation strategies include: the implementation of the legal binding frameworks for statuary key georegister and the European INSPIRE directive; supply optimization of governmental data, e.g., by creating a new Dutch SDI clearinghouse; chain cooperation to increase the use of geoinformation; and promotion of collaboration between government, businesses, and universities on innovation and economic value creation. The implementation process has been coordinated by the Ministry of Housing, Spatial Planning and Environment (now the Ministry of Infrastructure and Environment). The Geo-information Council, with representatives of all important governmental SDI stakeholders, is acting as the steering committee for the implementation of GIDEON. By the end of 2011, important progress was made in implementing GIDEON and development of the Dutch national SDI. However, as stated in the SDI monitoring report, not all implementation strategies have been fully executed and not all objectives have been reached (Geonovum 2011).

ANALYSIS
A wide variety of factors that determine the SDI collaboration process were mentioned in our interviews. On the basis of an analysis of the main interview themes, a distinction was made between the reasons why stakeholders were motivated to get involved in the collaboration process; critical factors that facilitated the SDI collaboration process; barriers that hampered collaboration between SDI stakeholders; and the dynamics of the collaboration process. Subsequently, for each theme, statements of the interviewees were abstracted. Table 2 summarizes the statements of the interviewees, together with the number of experts of the national SDIs of Spain (five interviews in total) and the Netherlands (six interviews in total) who made the statements.

The number of times statements were made by interviewees can be regarded as an indication of the importance of different factors and the way the collaboration process has evolved. They offer a point of departure for the analysis, but those statements depend very much on the context of each SDI and the overall viewpoint of the interviewee. Therefore, they needed to be studied in more detail. The following section further discusses the differences and similarities between the two case studies and elaborates on the statements made about collaboration factors and the dynamics of the collaboration process.

Table 2. Statements made and the number of times stated by the interviewees of the Spanish (SP) and Dutch (NL) national SDI, respectively

Motivation for SDI Collaboration:		SP	NL
1	Avoid duplication of efforts	3	1
2	Make information available for wider use	5	5
3	Better quality of information	0	1
3	Streamline information flows	4	5
4	Efficient information supply	2	5
5	Implementation of INSPIRE directive	3	5
6	Sharing of experiences and good practices	3	2
7	Creation of shared SDI facilities	2	3
Critical SDI Collaboration Factors:			
1	Good personal contacts	5	3
2	Getting the right people together	2	3
3	Knowledge exchange and discussions	3	5
4	Creating awareness	4	3
5	Attitude and engagement of stakeholders	5	6
6	Organizational/coordination structures (governance)	5	6
7	Legal frameworks	4	4
8	Shared vision and objectives	4	6
9	Definition of responsibilities and roles of stakeholders	2	2
10	A good business case (benefits should be clear)	0	4
11	Bottom-up approach	2	2
SDI Collaboration Barriers:			
1	Difficult to connect with users outside SDI community	2	4
2	Lack of time and resources	1	1
3	Technical and interoperability problems	1	2
4	Work depends on a few stakeholders	3	1
5	Complex administrative context/fragmentation	1	1
6	Difficult to give up competences and autonomy	2	3
7	Differences in SDI development between stakeholders	2	2
8	Unequal distribution of costs and benefits	0	1
SDI Collaboration Dynamic:			
1	Growth in SDI development	3	5
2	Ups and downs in collaboration	2	2
3	SDI is in a stable condition	1	0
4	Better organized and structured collaboration (governance)	4	6
5	More formalized collaboration (legal frameworks)	5	3
6	More intense collaboration between stakeholders	0	4
7	Involvement of more stakeholders	4	4
8	More coherent SDI strategy	1	4
9	Clarification of responsibilities and roles	3	3

MOTIVATIONS FOR SDI COLLABORATION

The major motivation for stakeholders to become involved in the development of both national SDIs was to streamline information flows and to make information available for wider use. Interviewees of both SDIs described the SDI concept as attractive for the effective management of geographic information supply in the public sector and for making geographic data and services more widely available. In Spain, stakeholders saw the necessity "to link information resources from different administrations" [SP5] and "making information of public administrations available and accessible" [SP2]. In the Netherlands, the development of the SDI should "streamline geographic information supply in the public domain" [NL1]. It was developed corresponding to the general concept of e-governance: "governmental (geographic) information should be easily available and accessible for everybody" [NL3].

Another main motivation for collaboration stated by interviewees from both SDIs is that individual organizations in the public administrations are lacking resources and knowledge to work independently on the implementation of SDIs and the creation of SDI facilities. This statement is much related to sharing of experiences and good practices, which also were stated by several interviewees of both SDIs. Interviewees consider the national SDI a good platform to help each other and collaborate on specific projects [SP5, NL5]. Furthermore, a main driver for SDI development and collaboration in both SDIs was the European INSPIRE directive. INSPIRE created a lot of awareness and exerted pressure from the "outside." In the Netherlands, "it forced stakeholders to take action and define together with other stakeholders their role in the SDI" [NL3]. In Spain, INSPIRE also created much awareness among Spanish stakeholders of the SDI concept and stressed the importance of participating in the development of the SDI [SP1].

Efficient information supply is mainly an important motivation for collaboration within the Dutch SDI. Interviewees stated that a main objective of the Dutch SDI is "to facilitate efficient geo-information exchange and supply" [NL1, NL3, NL6]. This statement is much related to "avoiding duplications of efforts," which is stated more by Spanish interviewees. They consider the INSPIRE principle of "data should be collected only once and kept where it can be maintained most effectively" as important motivation and guideline for collaboration with the Spanish SDI [SP1, SP2, SP4]. Only one interviewee of the Dutch SDI stated better quality of SDI information and services as motivation

for collaboration. Other interviewees didn't explicitly state it as important argument for their involvement in the SDI collaboration process.

Motivations for SDI collaboration stated by interviewees of both national SDIs are rather similar. Nevertheless, relevant differences can be noticed. Because of the structure of the Spanish government, decentralized in three main levels with a high level of responsibilities and self-government, collaborations within the Spanish national SDI have been focused on integrating regional/local initiatives. "Collaboration in the context of an SDI was seen as a solution for geographic Information sharing between different administrative levels in Spain [SP2]". Collaboration within the Dutch SDI has been more focused on an efficient management of geoinformation supply throughout the government. Main motivations of collaboration between Dutch stakeholders were efficiency gains and better performance of the Dutch government, in line with the e-governance concept.

CRITICAL SDI COLLABORATION FACTORS

The attitude and engagement of stakeholders with the SDI collaboration process have been stated by all interviewees as a critical factor in the development of their national SDI. Interviewees of both national SDIs stated that SDI development depends on good personal contacts and getting the right people together. In Spain, the working groups for the national SDI played an important role in facilitating collaboration by establishing personal relations and mutual understanding [SP3]. In the working groups, knowledge exchange took place and issues were discussed to further develop the SDI [SP1 and SP4]. In the Netherlands, there are no specific working groups for the development of the national SDI, but the Dutch SDI stakeholders "know each other and know their responsibilities" [NL3]. Collaboration is described as informal. There is "willingness and engagement to collaborate among stakeholders" [NL6].

Another critical factor stated by all interviewees of both national SDIs is the establishment of organizational and coordination structures that are supportive of collaboration. In Spain, all interviewees argued that establishing organizational structures was crucial for the SDI development and collaboration. The interviewees agree that the SDI legal framework approved in 2010, transposing INSPIRE, has been an important step toward better structured and more formalized collaboration. However, SP2 and SP5 argued that Spain is still facing difficulties to establish the required SDI coordination mechanism. In the Netherlands, the establishment of Geonovum and the Geographic Information Council has been, according to all interviewees, critical to better organize and structure SDI collaboration. This has been further supported by legal frameworks like INSPIRE and the core registers, which are "forcing different stakeholders to collaborate and therefore have played an important role in the SDI development" [NL2]. Furthermore, two interviewees of both national SDIs mentioned the importance of defining the responsibilities and roles of the stakeholders. "For good collaboration roles need to be clear; stakeholders need to have a common interest and need to be complementary" [NL5].

Having a shared vision and objective for the SDI development is seen as critical by almost all interviewees. For the Dutch SDI, the approval of the SDI vision and implementation strategy GIDEON in 2007 has been critical for the SDI development. GIDEON defined an ambition and objectives for different stakeholders stimulating the collaboration process [NL5, NL2, NL6]. GIDEON served as guidance for the development of the Dutch SDI and stimulated the involvement of stakeholders with the national SDI [NL4]. In Spain, SP1 and SP4 argued that the INSPIRE directive and its implementation worked like a strategic plan and gave a perspective for it and therefore has been critical. SP2 and SP5 also argue that a strategic plan is critical, but that Spain is lacking a strategic SDI plan defining objectives and priorities on the national level. Interviewees of both national SDIs also stated the importance of having a "bottom-up" approach for the development of the SDI. In Spain, geographic information duties and competences are distributed among many different stakeholders at distinct administrative levels. Therefore, Spain has used "a bottom-up approach; based on bringing SDI initiatives of distinct administrations together" [SP1]. Bottom-up development also is seen as important for the Dutch SDI, but in the Netherlands, it remains difficult to link the SDI to bottom-up developments in different application domains [NL1 and NL5].

The necessity to have a good business case and clear economic benefits was stated only by interviewees from the Dutch national SDI as being critical [NL4, NL5, NL3]. NL 3 argued that without "a good business case it is difficult for individual organizations to justify their investments in the development of the national SDI." Having a good business case was not mentioned by interviewees from the Spanish SDI.

Except for the statement of having a good business case, critical collaboration factors stated by interviewees of both national SDIs are rather similar. Main differences can be noticed in the way the national SDIs are implemented and collaboration has been organized. The development of the Spanish SDI has been focused on integrating regional/local initiatives with INSPIRE as important guidance. The collaboration process has been mainly based on consensus making starting without fixed regulations, with working groups serving as the germ and diffusion tool for the advance of the SDI in Spain. In the Netherlands, the development and implementation process of the SDI has been mainly based on two visionary documents describing the objectives and implementation strategy of the national SDI. From the beginning, attention has been given to establish the required organizational framework and to develop a coherent national SDI strategy.

SDI COLLABORATION BARRIERS

A barrier stated by most interviewees is the difficulty in connecting with the user community. As stated by NL1, "the application of data and services provided by the SDI is still limited. Only a

relatively small group of experts is using the Dutch SDIs." Also in Spain it remains difficult to connect and interact with users outside the SDI community [SP2, SP5]. According to NL4, for user communities it is still difficult to combine data and services from different sources for one purpose, because of a lack of knowledge. According to NL1, it requires "more interaction with broader user community groups, research communities, and citizens to stimulate and promote the use of high quality geographic data." Interviewees SP4 and NL4 argue also that the SDI community itself involves only a few stakeholders and that not enough resources are available to dedicate sufficient time and effort to work together on the further development of the national SDI. However, most interviewees did not state lack of financial and human resources as important barriers for SDI development.

Another barrier often mentioned is the difficulty faced by organizations in giving up competences and autonomy, e.g., in collecting and distributing specific datasets and developing their own services and tooling. In Spain, the sensitive relations between the national level and the regional level make prioritizing and political coordination of the SDI difficult [SP5]. Lower administrative levels are afraid to lose competences and higher administrative levels do not want to get involved in political problems. They want to maintain good relations with other administrations [SP3 and SP5]. This made it difficult to establish the required SDI coordination mechanism in Spain. Also for stakeholders in the Netherlands it was difficult to give up competences and autonomy on information supply [NL1 and NL3]. In the Netherlands, the geoinformation domain was fragmented with different stakeholders being greatly autonomous. "Stakeholders found it difficult to give up competences and autonomy on geoinformation information supply, which have a direct relation with their working processes" [NL3]. "It took a lot of effort to convince stakeholders that work could be done more efficiently with better quality if they would collaborate" [NL1].

Other barriers stated by interviewees were technical and interoperability problems and differences in SDI development between stakeholders. According to SP3, in Spain, many SDI initiatives in the regions and municipalities have been developed, but often they have been created independently from other initiatives. This has lead to differences and interoperability problems between them. According to NL4, within the Dutch SDI, not enough investments were made in creating the required technical facilities and, therefore, the technical infrastructure of the Dutch SDI is lagging behind. However, according to NL3, too often new technology is used that has not yet matured and thus is causing problems. Furthermore, several interviewees have mentioned differences between regions and municipalities in SDI activities as barriers for the development of the SDI on the national level. In Spain, there are differences between SDI developments in the different autonomous regions and it remains difficult to gain interest and participation of smaller municipalities and provincial councils [SP2 and SP5]. In the Netherlands, it remains difficult to get municipalities involved in collaboration processes for the development of the national SDI [NL2 and NL4].

Comparing the two national SDIs, barriers identified are rather similar. They seem mainly to be related to getting stakeholders involved and actively participating in the SDI development. A main challenge identified by the interviewees of both SDIs is to stimulate the application of data and services and the interaction with user communities. Interviewees of both SDIs also mention the complex administrative context as a barrier. In Spain, this is related to the decentralized structure of the government, with three main administrative levels with a high level of responsibilities. In the Netherlands, fragmentation of the geoinformation domain and the high level of autonomy of stakeholders have hampered the development of the national SDI.

SDI COLLABORATION DYNAMICS

Most interviewees stated that their SDI is gradually growing, but that the development of the SDI has its ups and downs. In Spain, most interviewees view the SDI as incrementally developing, but "that the process is often slowing down" [SP4]. Some moments there is a "strong peak, e.g., when new services are introduced, but in other moments not much progress is made" [SP1]. Only interviewee SP5 considers that at the moment the national SDI is in a stable condition for a while and has not developed lately. Most Dutch interviewees agree that the Dutch SDI is developing gradually, but not in a straight line. "Looking at the SDI development in the Netherlands, important progress has been made" [NL2]. "Different governmental institutes collaborate more on geoinformation issues and the SDI is improving" [NL5]. Interviewees also mention a tendency to involve more stakeholders in the collaboration process for SDI development. In Spain, SDIs with different administrative levels were developed and a network of contacts with representatives from different administrations has been built up [SP1 and SP5]. Within the Dutch SDI INSPIRE, the policy document GIDEON and the innovation program "Space for Geo-Information" played important roles in getting stakeholders involved. Space for Geo-Information stimulated "building a network and bringing stakeholders from academia, governmental institutes, and the private sector together" [NL5]. In the context of INSPIRE, "stakeholders get to know each other, which stimulated collaboration on other issues" [NL2].

A trend in both SDIs is the implementation of legal frameworks and the establishment of organizational structure for the SDI on the national level. In Spain, the SDI has been established as an infrastructure in which everybody could participate with not many obligations. However, with the approval of a first legal framework in 2007 and a second in 2010, SDI collaboration became more formalized. This gives the national SDI "a strong legal basis" [SP3 and SP4] and "means a change towards a more formalized approach" [SP1]. In the Netherlands, all interviewees agree that the organizational structure of the SDI has been improved by establishing Geonovum and the GI council. Furthermore, the establishment of legally binding frameworks, like the key registers and INSPIRE, contributed to improve SDI collaboration and at the same time have been the driving force for further SDI development [NL4 and NL5]. Interviewees of both SDIs stated that

those developments contributed to more coherent SDI strategies and a clarification of responsibilities and roles. In the Netherlands, the policy document GIDEON is seen as an important step to create a more coherent strategy linking the different SDI-related initiatives and clarifying responsibilities [NL2, NL3, NL5]. Also in Spain, the change to a more formalized approach is seen as an important step to rationalize collaboration within the national SDI by better defining responsibilities of the involved stakeholders [SP1 and SP3]. In general, interviewees identify an obvious trend toward formalization of the collaboration process. This is seen as important for the SDI development because it forced stakeholders to take action. However, the legislative frameworks are seen more often as "a ratification of informal collaboration practices that were taken place already" [SP2].

DISCUSSION

Our exploration of collaboration factors and dynamics within two SDIs identified a rather similar pattern. Also similar factors that determine the collaboration could be identified. Below we further discuss the main themes addressed in the interviews: motivation, critical factors, and barriers and dynamics of collaboration and compare our results with earlier research on SDI collaboration and highlight new findings.

MOTIVATIONS FOR SDI COLLABORATION

Main motivations for SDI collaboration identified in our case studies are: to streamline information flows, to make information available for wider use, and to make information supply more efficient. These motivations are in line with the findings of Nedović-Budić et al. (2004) and McDougall (2006). However, a motivation frequently mentioned in our study is the "implementation of the INSPIRE directive." There is an indication that obligatorily legal frameworks also can be important catalysts of SDI collaboration, which were not previously identified as important motivators for collaboration in previous work.

CRITICAL COLLABORATION FACTORS

Organizational structures and legal frameworks have been identified as two of the most critical collaboration factors. The other critical factor is the attitude and level of engagement of the stakeholders. This is in line with the findings of Tulloch and Harvey (2007) who concluded on the basis of a series of case studies "that most successful data-sharing networks relied on a combination of formal and informal relationships." Also, Nedović-Budić et al. (2004) concluded that formal mechanisms and informal interactions play significant roles in collaborative interorganizational data-sharing activities. A factor that is more eminent in our study compared to earlier work is the establishment of coordination and organizational structures of involved stakeholders: in the Netherlands, Geonovum and the GI council; in Spain, the SDI working groups.

SDI COLLABORATION BARRIERS

Main collaboration barriers were the difficulties for organizations to give up competences and autonomy and to establish relations with user communities. Also the challenge of SDI to attract users and to share competences to create integrated products and services also has been identified. These aspects are mentioned in previous studies by Díaz et al. (2011) and Nedović-Budić et al. (2008). However, little references could be found to empirical studies analyzing collaboration barrier within SDIs. Compared to the earlier work of Harvey (2001) and Nedović-Budić et al. (2004), technical issues and issues related to data access and standardized data exchange were less stated as critical for collaboration by our interviewees. Our findings, therefore, are in line with observations of Craglia et al. (2008) and Budhathoki et al. (2008), who stated that SDI implementation barriers are increasingly becoming nontechnical in nature.

SDI COLLABORATION DYNAMICS

Our results identified a dynamics of collaboration, in which more stakeholders are involved and a collaboration that is better organized and more formalized. This is confirmed by earlier SDI collaboration research. For example, Azad and Wiggins (1995), Kok van Loenen (2005), and Van Loenen and Van Rijn (2008) analyzed SDIs from an organizational perspective and identified an evolving dynamic in which interorganizational relations between SDI stakeholders pass from lower levels to higher levels of integration, when SDIs mature. The trend toward more formalized and agreed-on procedures also is identified by Craglia and Annoni (2007) and Lance et al. (2009). However, this development may be against the development of SDI collaborations based on self-organizing and more spontaneous interactions, as suggested by some other authors (Kok and Van Loenen 2005, Grus et al. 2010). Our results indicate that when SDIs are maturing, there is an increased need to define an organizational structure and discuss and apply a set of common procedures to manage and develop the SDI.

The dynamics of the SDI collaboration process in our study also have been identified in other studies on collaboration outside the SDI domain (Bailey and Koney 2000, Gadja 2004, Frey et al. 2006). Increasing levels of collaboration require increasing formalization and more specific definitions of roles and responsibilities (Todeva and Knokke 2005, Camarinha-Matos and Afsarmanesh 2007). Also, the level of engagement of stakeholders is identified by several studies as a crucial factor for collaborations (Todeva and Knoke 2005, Frey et al. 2006). When collaborations evolve, stakeholders are increasingly sharing ideas and knowledge to solve problems together (Camarinha-Matos and Afsarmanesh 2006). This tendency could be identified in our SDI case studies. A tendency that is less obvious in our research is the integration of operational activities and the creation of shared products and services. According to Camarinha-Matos and Afsarmanesh (2006), Allee (2008), and Navarette et al. (2010), one of the main purposes of collaboration is to integrate work processes and

create outcomes and value. In SDI collaborations, this remains challenging.

CONCLUSIONS

Our findings have provided insights into the factors and dynamics of collaboration in the SDI domain. Collaboration has been identified as critical for the development of SDIs. A pattern of an evolving collaboration with more stakeholders becoming involved and a collaboration process that becomes better organized and more formalized could be identified. Furthermore, a number of critical aspects and barriers have been identified that facilitate or inhibit the evolving collaboration process. For example, SDI collaboration requires both having formal legal and organizational structures in place and having good personal (informal) contacts and people engaged with the SDI concept and its benefits.

Our work contributes to earlier work on SDI stakeholder interaction and collaboration by giving more insights into the dynamics of the SDI collaboration process. It identifies a development of SDIs going beyond the data-sharing perspective, driven by technological and standardization issues. SDIs are getting more and more embedded in administrative organizational and legal structures. However, collaborations still are hampered by difficulties for organizations to give up competences and to establish relations with user communities. The main challenge for further development of SDI collaboration is, therefore, to develop structures where distributed competences and knowledge can be shared, enabling the creation of integrated products and services, with value for SDI users. This requires, as well, formalized collaboration structures as good informal contacts and engagement of involved stakeholders. In future research, collaboration processes and their evolution in time should be further examined and evaluated.

About the Authors

Watse Castelein is a researcher and Ph.D. candidate at the Technical University of Madrid in Spain and the Wageningen University in the Netherlands. His research interests include organizational and socioeconomic aspects of Geo-ICT developments with a specific focus on implementation and evaluation of spatial data infrastructures. As policy advisor, he contributed on the development and monitoring of the national Dutch and Spanish SDI strategy and the implementation of the European INSPIRE directive.

Corresponding Address:
MERCATOR Research Group
Technical University of Madrid (UPM)
Campus SUR, Paseo de la Arboleda s/n.
E-28031 Madrid (Spain)
E-mail: wcastelein@topografia.upm.es

Arnold K. Bregt is a professor of geoinformation science at the Wageningen University in the Netherlands. Following more than 20 years of experience in the field of geoinformation science research and applications, his current areas of interest

are spatial data infrastructures and sensors and human-space interactions. He has written for about 250 publications and holds an MSc and a Ph.D. from the Wageningen University.

Corresponding Address:
Centre for Geo-information
Wageningen University
Wageningen, The Netherlands
E-mail: arnold.bregt@wur.nl

Lukasz Grus is an assistant professor of geoinformation science at the Wageningen University. His current areas of interests are assessing spatial data infrastructures, evaluating the effects of open spatial data on organizations, and measuring how geographic information systems and spatial data infrastructures support the key processes of organizations. He has written for a number of publications in the GIS field and he holds an MSc and Ph.D. from the Wageningen University. He also works at Esri Netherlands as a business developer in a field of spatial data infrastructures.

Corresponding Address:
Centre for Geo-information
Wageningen University
Wageningen, The Netherlands
E-mail: lucas.grus@wur.nl

Acknowledgments

References

Akinyemi, F. O. 2011. Evaluating access to spatial data and information in Rwanda. Journal of the Urban and Regional Information Systems Association 23(2): 39-47.

Allee, V. 2008. Value network analysis and value conversion of tangible and intangible assets. Journal of Intellectual Capital 9(1): 5-24.

Andrade, A. D. 2009. Interpretive research aiming at theory building: Adopting and adapting the case study design. Qual Report 14(1): 42-60.

Ansell, C., and A. Gash. 2008. Collaborative governance in theory and practice. Journal of Public Administration Research and Theory 18(4): 543-71.

Azad, B., and L. L. Wiggins. 1995. Dynamics of inter-organizational geographic data sharing: a conceptual framework for research. In Onsrud, H. J., and G. Rushton. Eds., Sharing geographic information. New Brunswick, NJ: Center for Urban Policy Research, 2, 22-43.

Bailey, D., and K. Koney. 2000, Strategic alliances among health and human services organizations: From affiliations to consolidations. Thousand Oaks, CA: Sage.

Benbasat, I., D. K. Goldstein, and M. Mead. 1987. The case research strategy in studies of information systems. MIS Quarterly 11(3): 369-86.

Budhathoki, N. R., B. Bruce, and Z. Nedović-Budić. 2008. Reconceptualizing the role of the user of spatial data infrastructure. GeoJournal 72(3-4): 149-60.

Camarinha-Matos, L. M., and H. Afsarmanesh. 2006. Collaborative networks: Value creation in a knowledge society. In Wang, K., et al., Eds., Knowledge enterprise: Intelligent strategies in product design, manufacturing and management. Boston: Springer Publisher, 26-40.

Camarinha-Matos, L. M., and H. Afsarmanesh. 2007. A comprehensive modeling framework for collaborative networked organizations. Journal of Intelligent Manufacturing 18(5): 529-42.

Craglia, M., and A. Annoni. 2007. INSPIRE: An innovative approach to the development of spatial data infrastructure in Europe. In Onsrud, H., Ed., Research and theory in advancing spatial data infrastructure. Redlands: ESRI Press, 93-106.

Craglia, M., M. F. Goodchild, A. Annoni, G. Câmara, M. Gould, W. Kuhn, D. Mark, I. Masser, D. Maguire, S. Liang, and E. Parsons. 2008. Next-generation digital earth: A position paper from the Vespucci Initiative for the Advancement of Geographic Information Science. International Journal of Spatial Data Infrastructures Research 3: 146-67.

Crompvoets, J. 2006. National spatial data clearinghouses. Worldwide development and impact. Ph.D. Thesis. Wageningen: Wageningen University Press.

Daley, D. M. 2009. Interdisciplinary problems and agency boundaries: Exploring effective cross-agency collaboration. Journal of Public Administration Research and Theory 19(3): 477-93.

D'Amour D., M. Ferrada-Videla, L. San Martín-Rodríguez, and M. D. Beaulieu. 2005. Conceptual basis for interprofessional collaboration: Core concepts and theoretical frameworks. Journal of Interprofessional Care 19(1): 116-31.

De Andrade, F. G., C. d. S. Baptista, and F. L. Leite, Jr, 2011. Using federated catalogs to improve semantic integration among spatial data infrastructures. Transactions in GIS 15(5): 707-22.

De Man, W. H. E. 2013. Thinking outside the disciplinary box in coping with dilemmas in geoinformation management for public policy. Transactions in GIS 17(3): 452-62.

Diaz, L., C. Granell, M. Gould, and J. Huerta. 2011. Managing user-generated information in geospatial cyberinfrastructures. Future Generation Computer Systems 27(3): 304-14.

Elwood, S. 2008. Volunteered geographic information: Future research directions motivated by critical, participatory, and feminist GIS. Geojournal 72(3-4): 173-183.

European Commission. 2007. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Brussels: Commission of the European Communities.

Fletcher, B., W. Lehman, and H. Wexler. 2009. Measuring collaboration and integration activities in criminal justice and substance abuse treatment agencies. Drug and Alcohol Dependence 103(1): 54-64.

Frey, B. B., J. H. Lohmeier, S. W. Lee, and N. Tollefson. 2006. Measuring collaboration among grant partners. American Journal of Evaluation 27(3): 383-94.

Gajda, R. 2004. Utilizing collaboration theory to evaluate strategic alliances. The American Journal of Evaluation 25(1): 65-77.

Geonovum. 2011. Voortgangsrapportage uitvoering—monitoring GIDEON. Nummer 04, Amersfoort, The Netherlands.

Georgiadou Y., and J. Stoter. 2010. Studying the use of geoinformation in government—a conceptual framework. Computers, Environment and Urban Systems 34(1): 70-8.

Grus, L., J. Crompvoets, and A. K. Bregt. 2010. Spatial data infrastructures as complex adaptive systems. International Journal of Geographical Information Science 24(3): 439-60.

Harvey, F. 2001. Constructing GIS: Actor networks of collaboration. Journal of the Urban and Regional Information Systems Association 13(1): 29-37.

Jefatura del Estado. 2010. Ley 14/2010 de 5 de julio, sobre las infraestructuras y los servicios de información geográfica en España.

Klein, H. K., and M. D. Myer. 1999. A set of principles for conducting and evaluating interpretive field studies in information systems. MIS Quarterly 23(1): 67-94.

Koerten, H. 2011. Taming technology. The narrative anchor reconciling time, territory and technology in geoinformation infrastructures. Ph.D. Thesis. Delft: Delft University of Technology Press.

Kok, B., and B. van Loenen. 2005. How to assess the success of national spatial data infrastructures? Computers, Environment and Urban Systems 29(6): 699-717.

Lance, K. T., Y. Georgiadou, and A. K. Bregt. 2009. Cross-agency coordination in the shadow of hierarchy: "Joining up" government geospatial information systems. International Journal of Geographical Information Science 23(2): 249-69.

Masser, I. 1999. All shapes and sizes: The first generation of national spatial data infrastructures. International Journal of Geographical Information Science 13(1): 67-84.

Masser, I. 2007. Building European spatial data infrastructures. Redlands: ESRI Press.

Masser, I., A. Rajabifard, and I. Williamson. 2008. Spatially enabling governments through SDI implementation. International Journal of Geographical Information Science 22(1): 5-20.

McDougall, K. 2006. A local-state government spatial data sharing partnership model to facilitate SDI development. Ph.D. Thesis. The University of Melbourne.

Mezcua-Rodríquez, J. 2009. The spatial data infrastructure of Spain as an example of success in Europe. 9th United Nations Regional Cartographic Conference for the Americas, New York, August 10 to 14, 2009.

Navarrete, C., J. R. Gil-Garcia, S. Mellouli, T. A. Pardo, and J. Scholl. 2010. Multinational e-government collaboration, information sharing, and interoperability: An integrative model. In the proceedings of the 43rd Hawaii International Conference on System Sciences, Kauai, Hawaii, January 5 to 8, 2010.

Nedović-Budić, Z., and J. K. Pinto. 2000. Information sharing in an interorganizational GIS environment. Environment and Planning B: Planning and Design 27(3): 455-74.

Nedović-Budić, Z., J. K. Pinto, and L. Warnecke. 2004. GIS database development and exchange: Interaction mechanisms and motivations. Journal of the Urban and Regional Information System Association 16(1): 15-29.

Nedović-Budić, Z., J. K. Pinto, and N. R. Budhathoki. 2008. SDI effectiveness from the user perspective. In Crompvoets J., A. Rajabifard, B. Van Loenen, and T. A. Delgado Fernández, Eds., Multiview framework to assess SDIs. Melbourne: Melbourne University Press, 14, 273-303.

Neuvel, J. M. M. 2009. Geographical dimensions of risk management: The contribution of spatial planning and Geo-ICT to risk reduction. Ph.D. Thesis. Wageningen University.

Ødegard, A. 2006. Exploring perceptions of interprofessional collaboration in child mental health care. Journal of Integrated Care 20(6).

Onsrud, H. J. 1998. Compiled responses by questions for selected questions. Survey of national and regional spatial data infrastructure activity around the globe, Global Spatial Data Infrastructure. December 2010, http://www.spatial.maine.edu/~onsrud/GSDI.htm.

Paré, G., 2004. Investigating information systems with positivist case study research. Communications of the Association for Information Systems 13(1): 233-64.

Podolny, J. M., and K. L. Page. 1998. Network forms of organization. Annual Review of Sociology 24: 57-76.

Powell, W. W. 1990. Neither market nor hierarchy: Network forms of organization. Research in Organizational Behavior 12(6): 295-336.

Rajabifard, A., M. E. Feeney, and I. P. Williamson. 2002. Future directions for SDI development. International Journal of Applied Earth Observation and Geoinformation 4(1): 11-22.

Rajabifard, A., M. E. Feeney, I. Williamson, and I. Masser. 2003. National SDI initiatives. In Williamson, I., A. Rajabifard, and M. E. Feeney, Eds., Developing spatial data infrastructures: From concept to reality. London: Taylor and Francis, 6, 95-109.

SADL (Spatial Application Division Leuven). 2011a. Spatial data infrastructures in Spain: State of play 2011 (Leuven: K. U. Leuven + Margaret Hall, consultant).

SADL (Spatial Application Division Leuven). 2011b. Spatial data infrastructures in the Netherlands: State of play 2011 (Leuven: K. U. Leuven + Margaret Hall, consultant).

San Martin-Rodriguez L, M. D. Beaulieu, D. D'Amour, and M. Ferrada-Videla. 2005. The determinants of successful collaboration: A review of theoretical and case studies. Journal of Interprofessional Care 19(1): 132-47.

Thellufsen C., A. Rajabifard, S. Enemark, and I. Williamson. 2009. Awareness as a foundation for developing effective spatial data infrastructures. Land Use Policy 26(2): 254-61.

Todeva, E., and D. Knoke. 2005. Strategic alliances and models of collaboration. Management Decision 43(1): 123-48.

Tulloch, D., and F. Harvey. 2007. When data sharing becomes institutionalized: Best practices in local government geographic information relationships. Journal of the Urban and Regional Information Systems Association 19(2): 51-9.

Vandenbroucke, D., J. Crompvoets, G. Vancauwenberghe, E. Dessers, and J. Van Orshoven. 2009. A network perspective on spatial data infrastructures: Application to the subnational SDI of Flanders (Belgium). Transactions in GIS 13(1): 105-22.

Van Loenen, B., and B. Kok. 2002. Dutch NGII on course: A practical approach on implementing a vision. Proceedings of FIG XXII International Congress, Washington, DC, April 19 to 26, 2002.

Van Loenen, B., and E. van Rij. 2008. Assessment of spatial data infrastructures from an organizational perspective. In Crompvoets, J., A. Rajabifard, B. Van Loenen, and T. A. Delgado Fernández, Eds., Multi-view framework to assess SDIs. Melbourne: Melbourne University Press, 9, 173-92.

VROM (Volkshuisvesting Ruimtelijke Ordening en Milieu). 2008. GIDEON—key geoinformation facility for the Netherlands. Approach and implementation strategy (2008–2011). The Hague: VROM.

Warnest, M. 2005. A collaboration model for national spatial data infrastructure in federated countries. Ph.D. Thesis. The University of Melbourne.

Wehn de Montalvo, U. 2003. Mapping the determinants of spatial data sharing, Aldershot: Ashgate.

Yin, R. K. 2003. Case study research, design and methods, 3rd Ed. Beverly Hills, CA: Sage Publications.

Building the School Attendance Boundary Information System (SABINS): Collecting, Processing, and Modeling K to 12 Educational Geography

Salvatore Saporito, David Van Riper, and Ashwini Wakchaure

Abstract: *The School Attendance Boundary Information System (SABINS) is a social science data infrastructure project that assembles, processes, and distributes spatial data delineating K through 12th grade attendance boundaries for thousands of school districts in the United States. Until now, attendance boundary data have not been made readily available on a massive basis and in an easy-to-use format. SABINS removes these barriers by linking spatial data delineating attendance boundaries with tabular data that describe the demographic characteristics of populations living within those boundaries. This paper explains why a comprehensive GIS database of K through 12 attendance boundaries is valuable, how original spatial information delineating attendance boundaries is collected from local agencies, and techniques for modeling and storing the data so they provide maximum flexibility to the user community. The goal of this paper is to share the techniques used to assemble the SABINS database so that federal, state, and local agencies can apply a standard set of procedures and models as they gather data for their regions.*

INTRODUCTION

The School Attendance Boundary Information System (SABINS) is a spatial data infrastructure project that assembles and distributes kindergarten through 12th grade public attendance boundaries for thousands of school districts in the United States. School districts represented in the data collected during the SABINS project include more than half of all schoolchildren; these data are available free of charge from www.sabinsdata.org. Until the advent of SABINS, the Census Bureau's administrative units were the primary spatial data for researchers and policy makers interested in understanding a core social science issue: the impact of social context on life outcomes. An important goal of the SABINS project is to add to the quality of geographic and demographic data available to social scientists and policy makers who explore such issues. For example, researchers who study how neighborhood context influences educational outcomes, crime, disease, and related social processes typically must use areal units such as census tracts or block groups. While useful, these administrative geographies are limited for they fail to delineate socially meaningful boundaries that significantly affect the people who live, work, and play within them.

More particularly, they do not indicate which children have access to public educational facilities. Attendance boundaries—which are the catchment areas or zones drawn by local school districts to designate the housing units served by public schools—are spatial units that researchers can incorporate into studies that aim to improve the delivery of educational services. Yet, assembling and harmonizing attendance boundary geography for hundreds

of school districts is simply too expensive and time-consuming for small research teams and daunting for scholars whose expertise lies outside the domain of geographic information systems (GIS). Moreover, attendance boundaries present a variety of unexpected and difficult challenges, given the counterintuitive relationships between schools and the attendance boundaries they serve.

The SABINS project overcomes these challenges by creating a data structure that allows the seamless integration of attendance boundaries with three datasets: (1) school-level information from the National Center for Education Statistics Common Core of Data (CCD), which is a federal database describing the name, location, and student enrollment of all public schools in the United States; (2) complete count population data from the 2010 Census; and (3) detailed sociodemographic data from the American Community Survey (ACS).

The SABINS database contains kindergarten through 12th grade attendance boundaries for three states (Delaware, Minnesota, and Oregon), roughly 600 school districts embedded within 13 Metropolitan Statistical Areas (MSAs), and more than 400 of the largest school districts.[1] Figure 1 shows collected data in red. The goal of collecting sufficient data to cover half of the children in the United States was guided by several factors. Although there are more than 13,000 school districts in the United States, most children reside in roughly 800 of them. Thus, limited funding

---

[1] The metropolitan areas are Atlanta GA, Bakersfield CA, Hartford CT, Houston TX, Kansas City MO, Miami FL, Milwaukee WI, Minneapolis-St. Paul MN, Philadelphia PA, Portland OR, Orlando FL, Tampa FL, Tucson AZ, Virginia Beach VA, and Washington DC.
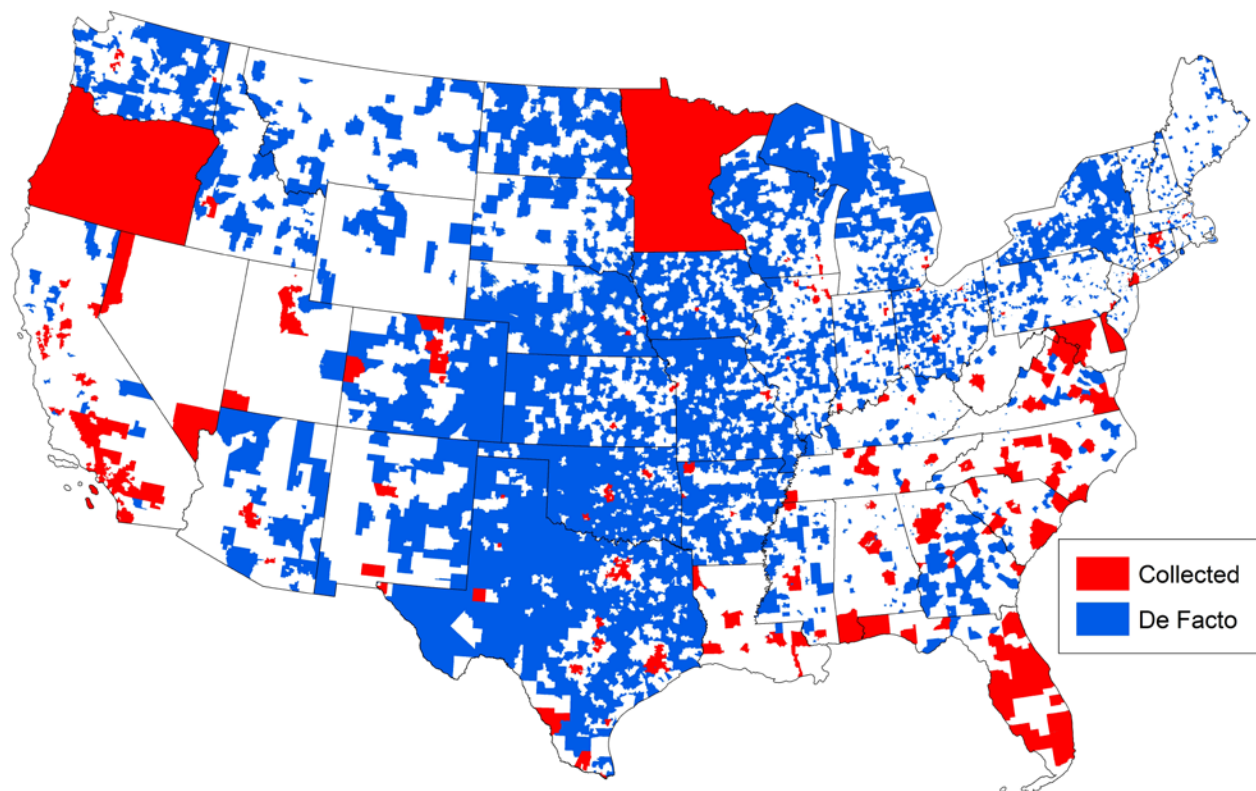
Figure 1.

produced a substantial database—more than enough information to allow researchers to undertake a variety of large-scale studies but at a cost much lower than necessary to collect data for all districts in the country.

At the same time, including districts embedded in metropolitan areas allowed the project to demonstrate the feasibility of collecting complete information from districts that varied in size—and thus to catalog the variety of ways in which local agencies maintained, stored, and distributed attendance boundary information. SABINS personnel collected data from more than 98 percent of the districts that it targeted. This success rate verified the feasibility of collecting data for the entire United States. The U.S. Department of Education has assumed data-collection efforts. It collected spatial data for the 2010–2011 to 2012–2013 school years for the 800 largest school districts and will expand the database to include the entire country for the 2013–2014 and 2015–2016 school years.

Additionally, the SABINS project has, for the first time, to the best of our knowledge, identified all school districts in the United States that are served by one and only one school in the district that has a kindergarten, a first grade, a second grade, and so on for every grade to grade 12. In these districts, every attendance boundary coincides with the entire school district boundary. These districts are de facto attendance boundaries since the Census Bureau distributes a national file of school district boundaries in GIS format. Nearly 60 percent of school districts in the United

States are de facto attendance boundaries and enroll roughly 20 percent of the public school students in the country. Identifying de facto attendance boundaries demonstrates the scope of work necessary for the U.S. Department of Education to expand data-collection efforts to include the entire country.

Another outcome of the SABINS project is to organize spatial data delineating attendance boundaries into a data model that provides users with the flexibility to analyze geographic areas that meet their particular needs. The SABINS project makes the following contributions. First, the data allows the extraction of *grade-specific* attendance boundary geographies for grades kindergarten through 12. If the goal is to build a national database of K to 12 educational geography, it is impossible to group attendance boundaries into a three-tier classification system of "elementary," "middle," and "high." The grade spans these labels signify vary widely across school districts. For example, a school district might have attendance boundaries that cover the typical K-5, 6-8, and 9-12 grade spans, but that same school district also may have boundaries that cover atypical grade spans such as K-2, 3-6, K-8, and 10-12. Other school districts have more than three geographic layers, some of which cover a single grade (e.g., grade 6).

The geographies in the SABINS database are integrated so that users can identify grade-specific boundaries that are geometrically coincident across specific grades. Users can deploy GIS software to identify attendance boundary polygons that are, for example, coincident for grades K, 1, 2, 3, 4, and 5. They then

can reconstruct this "K-5" boundary. Thus, if the attendance boundaries for grades K, 1, 2, 3, 4, and 5 coincide, these "K-5" boundaries can be assigned the same identification code for each of these grades. Our database allows users the flexibility of working with grade-specific boundaries or boundaries that span grades, allowing users to define "elementary," "middle," and "high" schools as they see fit.

The SABINS project also has created a data model that specifies the "many to many" relationship between attendance boundaries and the schools that supply services to them. Schools and their corresponding boundaries are related but are not equivalent. While most schools serve one boundary, there are deviations from this dominant pattern, including: (1) two or more schools can provide services to the same boundary; (2) two or more schools provide services to a portion of an overlapping boundary; (3) the same school provides services to different boundaries at different grade levels—e.g., a school serves a kindergarten boundary that covers a different area than the first grade boundary; (4) the same boundary is served by different schools at different grade levels; and (5) a school supplies services to attendance boundaries located in different school districts (or even in different states).

The final outcome of the SABINS project is to integrate attendance boundaries with census geography. Every attendance boundary is associated with census blocks. This relationship facilitates the summary of block-level population characteristics to: (1) grade-specific attendance boundaries; (2) attendance boundaries that coincide across grade spans; (3) schools that provide services to specific areas. Finally, the SABINS project integrates attendance boundaries with detailed sociodemographic data from the Census Bureau American Community Survey. American Community Survey data are summarized to block groups—but block groups do not nest within attendance boundaries. To overcome the misalignment between these geographies, the SABINS project uses a straightforward spatial-allocation technique to estimate detailed population characteristics within attendance boundaries.

THE USEFULNESS OF ATTENDANCE BOUNDARIES

Some scholars argue that census tracts and other administrative geography (e.g., block groups and zip codes) are questionable proxies for neighborhoods (Sampson et al. 2002) and are clearly not interchangeable with attendance boundaries themselves. Yet, because of a lack of school boundary geography, researchers are forced to make the simplifying assumption that the census tract inside of which a school is located is an adequate proxy for its attendance boundary (Card and Krueger 1992, Entwisle et al. 1997, Reardon and Yun 2001, Frankenburg et al. 2003, Logan and Oakley 2004, Owens 2010). One problem with this assumption is that the population characteristics of attendance boundaries are imperfectly correlated with those of their proxy areas.

To explore this issue, we examine data for Delaware. As shown in Table 1, the correlation coefficient between the percentage of non-Hispanic black people in census tracts in which schools

are located and the percentage of non-Hispanic black people in schools' actual attendance boundaries is .87; this correlation is .41 for grade 7 boundaries and .58 for grade 12 boundaries. Similar correlations exist between block groups and attendance boundaries. The data also are consistent for other racial groups. Beyond the imperfect correlations between actual and proxy zones, certain census tracts or block groups do not contain a school while others contain multiple schools. Using proxy zones results in counting populations in some areas multiple times while failing to include populations in areas that do not contain a school. Thus, using tracts and block groups leads to inaccuracies when tabulating population totals for an entire school district.

Table 1. Correlation coefficients between neighborhood racial composition across geographies by grade levels

Percent Non-Hispanic African-American			
Kindergarten	(1)	(2)	(3)
(1) School Boundary	1.00		
(2) Census Tracts	0.87	1.00	
(3) Block Groups	0.82	0.94	1.00
Grade 7	(1)	(2)	(3)
(1) School Boundary	1.00		
(2) Census Tracts	0.41	1.00	
(3) Block Groups	0.34	0.90	1.00
Grade 12	(1)	(2)	(3)
(1) School Boundary	1.00		
(2) Census Tracts	0.58	1.00	
(3) Block Groups	0.66	0.92	1.00
Number of kindergarten observations is 84; grade 7 is 33, and grade 12 is 23.			

Attendance boundaries provide a more accurate estimate of the population characteristics of students who may be enrolled in a school than does the census tract in which a school is located. Still, not all students who live within attendance boundaries attend their local public schools. In the United States, as in many other countries, parents with sufficient economic means can send their children to private schools. Another alternative to private schools are magnet schools. While every district that has a magnet school program implements distinct magnet school policies, most magnet school programs have similar features: They allow any student within a district to apply to a magnet school, are not tuition-based, and only grant admission if a student applies. Student applications may be approved or denied based on several factors, including lottery, standardized test scores, and previous academic performance. Charter schools are similar to magnet schools for they are publicly funded and typically do not have defined attendance boundaries.

As shown in Table 2, during 2010, 10.7 percent of students attended private schools and 5.7 percent attended magnet or char-

Table 2. Percent of students enrolled in public school by grade span, 2009

	Kindergarten	1 to 4	5 to 8	9 to 12	Total
% in Private	13.2	10.7	10.8	9.9	10.7
% in Magnet	2.0	2.1	2.5	3.8	2.8
% in Charter	3.2	3.0	3.0	2.7	2.9
Total	18.4	15.8	16.3	16.4	16.4

Sources: Percent public school enrollment from the 2009 American Community Survey; percent magnet and charter school enrollment derived from 2009–2010 school year Common Core of Data (Chen 2011).

ter schools in the United States. It is important to emphasize that public schools provide educational services to a fixed area—and it is possible to determine the characteristics of students who live in the area a school serves—but this information is an imperfect indicator of the characteristics of children who are actually enrolled in a school. As we describe in greater detail below, it is possible to estimate the number of students in an attendance boundary who are enrolled in a private school. This information can be used to assess the extent to which the demographic compositions of schools' students are similar to those in their corresponding catchment areas. The SABINS data also can be combined with the Common Core of Data to estimate the number of students in a district who are enrolled in magnet and charter schools (i.e., those schools that do not serve a fixed catchment area).

It is much more of a challenge to determine the number of *public* students who live in a school catchment area but who are not enrolled in the school that serves their residence. Some school districts have various transfer programs that allow students to attend a school of their choice. In some urban school districts, public school students enroll in charter and magnet schools at relatively high rates. It is more difficult to create an estimate of the number of children in an attendance boundary, but, with a few assumptions, this is possible as well.[2]

Despite these limitations, three important points must be emphasized: First, the American Community Survey data distinguish between public school and private school children, making it possible to create demographic profiles of attendance boundaries for *public school children.* Researchers already have used the SABINS data to distinguish between public school and private school students who live within an attendance boundary (National Research Council 2012.) Second, although attendance

boundaries are "permeable," they are much better than substituting census tracts, block groups, and zip codes as proxy boundaries. Third, educational geography is useful in and of itself. For example, scholars and policy makers can use attendance boundary data to investigate unique questions such as the extent to which local districts delineate their boundaries to reduce or contribute to racial and economic segregation (Heckman and Taylor 1969), for studies of public health and epidemiology (Elliott and Wartenberg 2004, Diez Roux 2001, Krieger 2006, Krieger et al. 2002, Shai 2006, Winkleby and Cubbin 2003, Xue et al. 2009), to study the effects of school quality on housing values (Black 1999, Brunner et al. 2002, Brunner et al. 2001, Downes and Zabel 2002, Ioannides 2004, Weimer and Wolkoff 2001), and to understand the factors that lead to public and private school choice (Saporito 2009). Specific policy and planning applications using attendance boundaries include building safe walking and biking routes to school (Huang and Hawley 2009); estimating public school populations eligible for subsidized school meals; school enrollment projections (Edwards and Ehrenthal 2008); and efficient bus routing and siting new school construction (Lemberga and Church 2000).

BUILDING THE SCHOOL ATTENDANCE BOUNDARY INFORMATION SYSTEM

One of the most challenging tasks in building the SABINS database was collecting the source information from hundreds of local school districts and county GIS offices. One goal of this paper is to document the feasibility of collecting and compiling this information. The very largest school districts—those that enroll 20,000 or more students—have scores or hundreds of K-12 attendance boundaries, but this information exists in a wide variety of formats. Most school districts post attendance boundary information on their Web pages to inform parents about which schools their children should attend. This information is typically *displayed or described* in one of four formats: (1) static, cartographic images (e.g., PDF images) displaying attendance boundaries and the streets and other line features that attendance boundaries follow; (2) interactive, Web-enabled maps that allow parents to pan and zoom to areas within a school district; (3) narrative or legal descriptions that verbally describe the boundaries; and (4)

---

[2]   The SABINS database can be used to identify "neighborhood" schools that serve a fixed catchment area. Any school that is part of a school district but is not in the SABINS database is a non-neighborhood school (and most of these are classified as charter and magnet schools). Users can combine the SABINS database with the CCD and create their own estimates. As an example, there could be a situation in which 500 first grade students live in an attendance boundary. Of these first graders, 400 are enrolled in public schools. And, there are 300 first grade students in the school that serves the attendance boundary. It is reasonable to assume that 100 first grade students who live in the attendance boundary are enrolled in non-neighborhood public schools.

Web-enabled "address locators" that allow parents to enter their residential addresses into a search engine. All this information is used to digitize attendance boundaries. If a school district does not display maps of its attendance boundaries, the maps usually can be obtained by filing a public information request with a school district.

DIGITIZING PROCEDURES

The SABINS project digitizes "analog" information such as static images of attendance zones, narrative/legal descriptions of attendance boundaries, or lists of addresses that schools serve. The primary base layer used to digitize attendance boundaries includes the line features from the Census Bureau 2010 Topologically Integrated Geographic Encoding and Referencing System (TIGER/Line files). The advantages of using TIGER/Lines are their ready availability and metadata that document the positional accuracy of its features. To ensure that the accuracy of boundaries is preserved, line features that make up the outlines of an attendance boundary are "geotraced," and thus attendance boundaries adopt all the vertices of the TIGER/Line features.

Although most attendance boundaries follow line features that are represented in the TIGER/Line files, certain portions of some boundaries do not. This occurs, for example, when a portion of an attendance boundary encompasses addresses on both sides of a street. In such cases, we obtain spatial data from Esri's "ArcGIS online" database,[3] which has two important data sources. The first database is the Digital Orthophoto Quarter Quad (DOQQ) areal imagery (at a resolution of one meter), which identifies features of interest (typically housing units). A second source of information is parcel data. Many counties make their data available via ArcGIS Online and, when available, these actual parcel or cadastral data are used as a base layer for digitizing. These two resources are used to ensure that attendance boundary polygons do not cut across housing units.

Although these digitizing methods produce a set of electronic GIS files that follow a consistent system, it is preferable to obtain original electronic GIS files from a school district. Many school districts create digital GIS files in-house or have them made by a consulting firm. School districts usually share these data upon

3   A University site licenses allows the SABINS project to digitize boundaries using the DOQQ imagery provided by Esri and still maintain oversight of the attendance boundary vector data that are produced. For those who do not have access to DOQQ imagery via an Esri site license, the DOQQ imagery can be obtained directly from the U.S. Geological Survey.
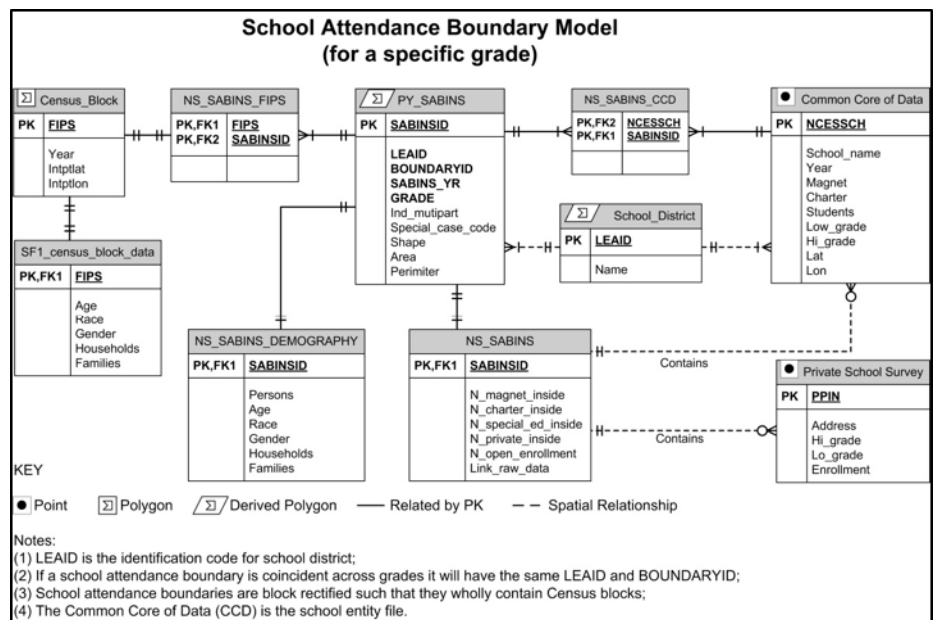
Figure 2. Logical Data Model Representing Relationships Among School Boundary Entities.
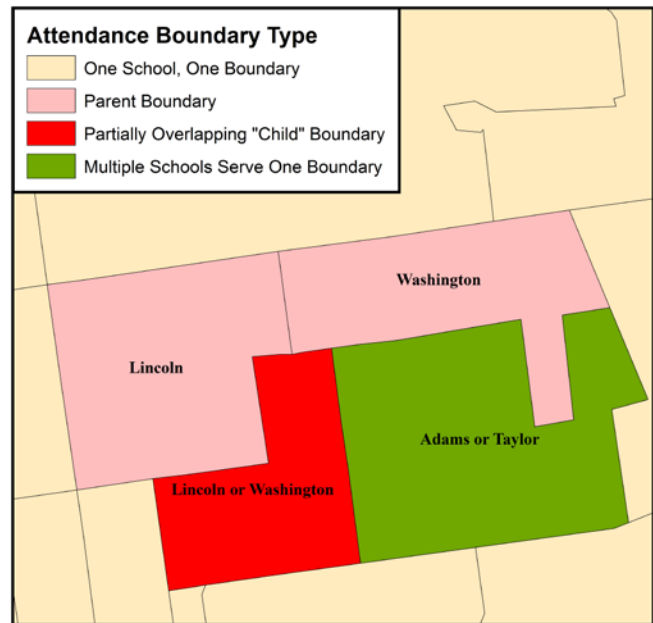
Figure 3. Types of Attendance Boundaries.

request—typically as Esri shapefiles or geodatabases, MapInfo files, or Computer Aided Drafting files. The quality of these spatial data varies widely. In best cases, cadastral data were used to digitize boundaries. In other cases, the quality of the digital GIS files is poor—boundaries do not have or enforce topology or do not carefully follow visible line features, and include many gaps and overlaps in a layer. Despite the lower quality of some GIS files obtained from local agencies, these data often are the only information available.

LOGICAL GIS DATA MODEL

The primary entities in the SABINS database consist of at-

tendance boundaries, the public school or schools that provide educational services to each attendance boundary, and the census blocks that lie within each attendance boundary (see Figure 2). Related entities include the school districts that contain attendance boundaries and the number of private, charter, and magnet schools located within attendance boundaries.

It is tempting to think of schools and attendance boundaries as the same entity, and many local districts simply assign a school's identification code—usually in the form of a school name—to the boundary to which it supplies services. However, schools and their corresponding boundaries do not have a one-to-one relationship. As shown in Figure 3, some noteworthy relationships in the model include:

(1) Two or more schools that provide services to the same attendance boundary. This scenario is depicted by the green polygon. Children who live within the green-shaded attendance boundary attend either Adams or Taylor.

(2) Two or more schools that provide services to an overlapping portion of two or more "parent" attendance boundaries. This scenario is depicted by the light red boundaries (i.e., the parent boundaries) and the dark red boundary (i.e., the partially overlapping "child" boundary). Children who live in the dark red polygon have the option of attending either Washington or Lincoln. Children who live in the light red boundary labeled "Washington" must attend Washington school and children who live in the boundary labeled "Lincoln" must attend Lincoln school.

(3) One school can serve two or more areas at different grade levels. For example, the *same school* can serve one attendance boundary for grades kindergarten to 5 and another attendance boundary for grades 6 to 8. Cities such as New York, Chicago, Los Angeles, Philadelphia, and Detroit all have instances of multiple boundaries for single schools.

(4) Multiple schools can serve the same attendance zone at different grade levels. For example, the same attendance zone can be served by a school that has grades kindergarten to 2. That same attendance zone can be served by a different school that has grades 3 to 5.

(5) A school can provide services to attendance boundaries in different *school districts*. This scenario typically occurs in rural areas in the higher grade (typically, grades 9 to 12). In this scenario, the school district's polygon is preserved, but the school that provides services to both attendance boundaries is associated with each polygon.

These relationships are counterintuitive to the commonplace notion that one school serves one attendance boundary, or that every school only serves children within a single school district. To accommodate these relationships, attendance boundaries are assigned identification codes that are separate from but linked with school identification codes. Unique school identification codes for all 103,000 schools in the United States are obtained from the Common Core of Data (U.S. Department of Education 2010). A relational table links schools in the Common Core of

Data with attendance boundaries in the geographic file.

The related tables address situations (1) through (5) above. Situation (1) is addressed in a straightforward manner by building a relational table that links schools and attendance boundaries. A public school student who lives in one of these "optional" attendance boundaries can chose to enroll in one of the neighborhood schools that supplies services to the area. Situation (2) occurs when multiple schools serve an "overlapping" area—as shown by the polygon shaded in dark red in Figure 3. In this scenario, the "overlapping" portion of the attendance boundaries served by Lincoln and Washington is treated as a separate polygon with a unique identification code. If two or more public schools provide educational services to an area, the children who live there can select one of the schools that supply services to it. The third situation (3) occurs when the same school has different boundaries at different grade levels. In this case, each attendance boundary is assigned a unique identification code. Situation (4) arises when the same attendance zone is served by different schools at different grade levels; again, this is modeled by assigning the attendance zone a unique identification code and linking it with the schools that serve it. Situation (5) occurs when a school supplies services to children in an attendance boundary in different school districts. The portion of the attendance boundary that lies within each school district is assigned a unique boundary ID. As shown in the logical data model, these situations result in a "many to many" relationship between attendance boundaries and schools.

CREATING GRADE-SPECIFIC GEOGRAPHY

While the entities in the SABINS data model are few, the model masks some of the complex spatial and tabular relationships among these entities. In particular, attendance boundaries are typically thought of as three layers consisting of "elementary," "middle," and "high" school polygons. The terms *elementary, middle,* and *high* have no standard grade ranges and are merely convenient labels for attempting to describe schools that provide services to "youngsters," "adolescents," and "teenagers." Indeed, there are 91 possible grade spans that an attendance boundary can cover (e.g., grade K, grades K to 1, grades K to 2, and so on to grade 12) and there are attendance boundaries that cover most of these 91 possible grade-span combinations.

To overcome this challenge, the SABINS database contains attendance boundaries by grade level. Users have access to grade-specific boundaries or boundaries that can be reassembled to cover grade spans.[4] A simple example illustrates the challenges presented when trying to configure "elementary" attendance boundaries. There are many cases in which some sixth grade attendance boundaries are embedded in an "elementary" layer

---

4   The SABINS project preserves all of the original "elementary," "middle" and "high" attendance boundaries. Some school districts have four or five sets of boundaries, with labels such as "primary" "intermediate," or "junior high." Other districts have two layers.

while other sixth grade boundaries are embedded in the "middle" layer. If a school district represents its sixth grade attendance boundaries in separate layers, it is not possible to determine the school assignment of all sixth grade students by examining only the elementary school geography. In this situation, it is necessary to create a separate attendance boundary layer for sixth grade by merging the sixth grade polygons from the elementary and middle layers. This principle holds true for all grades. Thus, the SABINS project creates 13 geographic layers, one for each grade K through 12.[5] Moreover, some school districts have "sixth grade attendance boundaries" in addition to their elementary, middle, and high school layers.

While it is true that every area within a school district must be covered by an attendance boundary for each grade, many attendance boundaries coincide across grade spans. Indeed, in almost all school districts, the second grade attendance boundaries coincide with third grade attendance boundaries. It is valuable to know whether boundaries coincide across grade levels and this information is ideally preserved in the primary key of each attendance boundary polygon. The primary key consists of four fields: (1) LEAID, which is a unique identification code for every school district and is derived from the U.S. Census Bureau; (2) BOUNDARYID, which is the unique identification code for every attendance boundary within a district; (3) YEAR is the school year for the data (where a value of "10" is the 2009–2010 school year); and (4) GRADE, where a value of "00" represents kindergarten. If an attendance boundary is the same across grades (for the same year), the values for the fields LEAID and the BOUNDARYID are the same. For example, if all the attendance boundaries in a given district coincide for grades K through 5, then values for LEAID, BOUNDARYID, and YEAR will be the same.

PROCESSING STEPS TO IMPLEMENT THE DATA MODEL

When GIS data of attendance boundaries are available from local agencies, it is preferable to process these files rather than digitize the attendance boundaries from paper maps or narrative descriptions. The primary advantage of obtaining existing GIS data is that it saves time (particularly for the largest school districts). Yet, digital GIS data from local agencies often have the following deficiencies: (1) there are no topological rules established or enforced in the geography and some school districts may have hundreds of gaps and overlaps between their attendance boundaries; (2) some portions of "elementary," "middle," and "high" attendance boundaries should share the same line segments but often do not. As discussed below, this inconsistency can cause problems in associating census blocks to attendance boundaries

across grade-specific boundaries; (3) school districts need to be edge-matched to eliminate gaps and overlaps between them; (4) most attendance boundaries do not have an identification code that is distinct from the school or schools that provide services to the attendance boundary; (5) multipart polygons typically are treated as single-part polygons; and (6) the files do not necessarily identify the grades that a particular attendance boundary serves. Despite these shortcomings, it still is preferable to obtain and "clean" the digitized and attributed GIS data rather than digitize boundaries from paper records. It also is easier to process data that are digitized by SABINS staff—even though these data do not have the deficiencies listed previously.

The SABINS project has written a series of custom GIS programs that quickly, consistently, and accurately correct the problems associated with attendance boundary geography. Indeed, these scripts are used to process all GIS data, whether or not they were digitized by SABINS staff or were obtained directly from local agencies. The first processing step consists of assigning all attendance boundaries a common set of fields in the attribute table. All attendance boundaries are assigned a field called "source name" that contains the "identification" code that school districts assign each attendance boundary; most school districts identified their attendance boundary with a single field that contains the name(s) of the school(s) that provide services to an attendance boundary. This "source name" was preserved throughout all processing steps as a means of quality assurance. Any single-part polygons that are supposed to be the same—as indicated by the same source name—are dissolved into multipart polygons.

Because local agencies typically conflate schools with attendance boundaries, the second processing step is to assign each attendance boundary an identification code that is preserved throughout all processing stages. This identification code preserves the original geography obtained from school districts. This identification code is created by concatenating three separate fields. The first is a school district identification code number, called the Local Education Agency ID (or LEAID). The second field is a school level field, where the character "E" is assigned to elementary school polygons, while the characters "M" and "H" are assigned to middle and high attendance boundaries, respectively. (When a school district has, for example, five layers, including "primary" and "intermediate" layers, they are labeled with a "P" and "I," respectively.) The third field is a sequential set of numbers that are automatically generated in the attribute table of ArcMap 10's shapefiles and feature classes (and stored as a static field in the attribute table). Concatenating these three fields allows SABINS to reproduce the original "primary," "elementary," "intermediate," "middle," and "high" school polygon layers that were delineated by local school districts—but this identification code distinguishes these original boundaries from the schools that supply services to them. SABINS staff members also assign attributes to the original input layers that identify the lowest and highest grades that an *attendance boundary* serves. These attributes preserve the original grade span of the input

---

[5]   A school may have different boundaries at different grade levels. For example, a school's K to 5 boundaries cover a different area than its 6 to 8 boundaries. This is one of the reasons why we provide grade-specific boundaries rather than assigning each attendance boundary the same identification code as the school.

boundaries and allow the creation of grade-specific attendance boundaries and the original input boundaries.[6]

ASSOCIATING SCHOOLS WITH ATTENDANCE BOUNDARY POLYGONS

Attendance boundaries then are assigned a series of fields that contain the unique identification codes of the schools that provide services to them. The school identification codes are derived from named "NCESSCH" in the CCD. The process of assigning the NCESSCH school identification codes to every polygon is completed with a "fuzzy string" name-matching algorithm.[7] This algorithm links the name of the school that serves an attendance boundary (which is typically in the attribute table of a GIS layer) and finds its corresponding school name in the Common Core of Data. For example, the polygon layer may have the school name "John P. Jones" in the GIS attribute table, while the CCD could have the school name "John Paul Jones Elementary School." The NCESSCH identification code from the CCD is populated in the attribute table of an attendance boundary polygon layer by using the custom algorithm. (The school name from the CCD also is populated in a separate field of the attribute table to ensure the accuracy of the school assignment.) If two or more schools serve an attendance boundary, their corresponding NCESSCH identification codes are stored in a subsequent series of separate fields. As a quality-assurance measure, the name-matching algorithm also determines if a school is inside the attendance boundary it serves.

Each attendance boundary then is assigned 13 fields for grades kindergarten through 12. If an attendance boundary is contained within an "elementary layer," the fields are named "E_00" to "E_12"; middle school layers are given field names of "M_00" to "M_12"; and high school layers are given "H_00" to "H_12." These fields are set to null initially. Each unique "elementary," "middle," and "high" attendance boundary then is joined with the CCD (by the NCESSCH code). If the input *attendance boundary* indicates that it serves grades 00 to 05, then the attendance boundary is assigned the proper NCESSCH code for those grades. Specifically, if an "elementary" attendance boundary serves grades 00 through 05, the E_00 through E_05 fields are

Table 3. Assignment of grade-specific school identification codes to the "elementary" school layer

Name of Grade-specific Field	NCESSCH Codes
e_00	1200200001
e_01	1200200001
e_02	1200200001
e_03	1200200001
e_04	1200200001,1200200002
e_05	1200200001,1200200002
e_06	1200200001,1200200002
e_07	Null
e_08	Null
e_09	Null
e_10	Null
e_11	Null
e_12	Null

assigned appropriate NCESSCH codes,[8] while the E_06 through E_12 fields remain null. This is accomplished with custom python scripts. A typical middle school will have NCESSCH codes for the M_06 through M_08 fields and the remaining fields remain null; similarly, a typical high school will have values for grades 9 through 12.

If a grade-specific boundary is served by more than one school, then the NCESSCH identification codes of all schools that serve the boundary are stored in a comma delimited list (where the identification codes are sorted from low to high). For example, an attendance boundary in the elementary layer can be served by two schools: Jones School where NCESSCH is "1200200001" and Montgomery School where NCESSCH is "1200200002." The CCD indicates that Jones School offers grades 00 to 06 and Montgomery School offers grades 04 to 06. This results in populating grade-specific fields, as shown in Table 3.

STORING THE SEPARATE LAYERS AS A UNION

Once each elementary, middle, and high school attendance

---

[6] It is necessary to record the lowest and highest grades that an *attendance boundary* serves since a single school can serve multiple boundaries at multiple grade levels. As noted previously, a school can serve two boundaries (e.g., one boundary for grades Kindergarten to five and another boundary for grades six to eight). To create grade-specific boundaries—and to avoid falsifying data—it is necessary to identify the lowest and highest grade a unique attendance boundary serves. But assigning identification codes to the unique input geographies also has that advantage of allowing users to obtain the original "elementary," "middle," and "high school" layers obtained from school districts.

[7] As a quality assurance measure, the algorithm also insures that all schools in the CCD (for a particular school district) are matched with a record in the attribute table of the attendance boundary polygons.

[8] This processes uses information describing the grade range of the *boundary* and the grade range of the *school*. For example, the boundary could cover grades Kindergarten to five. Yet two schools can serve that boundary. Each school could have separate grade ranges. The first school will range from grades Kindergarten to two and the second from three to *eight*. As a consequence, the school identification codes assigned to the boundary will cover Kindergarten to five, but the identification codes for the first school will start at Kindergarten and end at the second grade and the second school will start and the third grade and end at the *fifth grade (not the eighth grade)*. The heart of the SABINS project is that it treats boundaries and schools as separate entities.

boundary is linked with appropriate school identification codes, then a unioned feature class is created. This process results in creating a new set of polygons (stored in a single layer) that represents the unique intersections among the polygons from the original layers. This overlay process quickly eliminates gaps and overlaps between polygons (or holes within polygons) and conflates line features from the elementary, middle, and high school layers. The conflation process ensures that lines that should be the same are the same.

The SABINS project uses a tolerance level of no more than 30 feet—and frequently uses a tolerance setting of ten feet.[9] The reason we adopt a maximum threshold of 30 feet is that this length is about the width of a small tertiary street or a small alley; moving vertices that accurately follow the centerlines of narrow roads do not degrade the accuracy of the source data. The union process is essential for enforcing topological rules. Correcting gaps and overlaps in the original elementary, middle, and high school attendance polygon layers is not sufficient. A simple example illustrates the challenge. If some sixth grade attendance boundaries are represented in the "elementary" school layer and others are represented in the "middle" and "high" school layers, creating a single sixth grade file requires merging boundaries from all three layers. Thus, it is not sufficient to enforce topology for each of the three layers separately and then merge the sixth grade polygons that originated from each layer. Simply merging the sixth grade polygons that were derived from the (topologically corrected) elementary, middle, and high school layers still will lead to gaps and overlaps between the sixth grade polygons that originated from the original three layers. The union process eliminates this problem by conflating line features from the three layers.

Although the union immediately eliminates gaps, overlaps, and small holes within polygons in the same layer—and conflates line features across layers—a union among the three layers often creates new "sliver" polygons. These are eliminated using the following steps. Multipart polygons are "exploded" into single-part polygons. Any sliver that is not covered by all the original "elementary," "middle," or "high" school layers are identified and selected automatically. "Small" slivers (i.e., those less than 10,000 square feet) also are automatically selected. Selected slivers are merged with adjacent polygons.

In addition to enforcing topology quickly, the union of the layers ensures that the entire area within a school district (e.g., all intersections among "elementary," "middle," and "high" school polygons) is covered by every grade K through 12. Custom GIS scripts ensure that every square inch of a school district is served by a K to 12th grade boundary and that every boundary is served by at least one school for every grade. This ensures that students at all grades know the boundaries in which they live and the schools

that provide services to them.

The final step in the union process is to add 13 new fields to the attribute table. These fields store the NCESSCH identification codes of the schools that provide services to each attendance boundary. (The NCESSCH identification codes for a grade are sorted from low to high within a single field and are comma delimited.) These 13 fields are "final_id00" to "final_id12." The code that undertakes this task determines if, for example, a fifth grade attendance boundary polygon originated from the elementary, middle, or high school attendance boundary file. The logic of the code can be summarized as follows: If E_05 has an NCESSCH code, then the "final_ID05" field is assigned the NCESSCH school identification code from the elementary school layer; if M05 has an NCESSCH code, then the "final_ID05" field is assigned the NCESSCH code from the middle school layer. Once all 13 final_ID fields are assigned NCESSCH codes, the union is dissolved 13 times—one time for each field final_ID00 to final_ID12. This entire process creates 13 topologically correct and geometrically consistent grade-separated polygon layers, each of which has an attribute describing the school or schools that supply services to a grade-specific attendance boundary. Each of the boundaries in the 13 feature datasets then is assigned a permanent identification code (i.e., the SABINSID as shown in the PY_SABINS feature dataset in Figure 2). At this stage, the "NS_SABINS_CCD" associational table is created in order to follow the normal forms necessary for robust spatial database management systems.[10] One final processing step incorporates ancillary information about attendance boundaries. This information is stored in the "NS_SABINS" table shown in Figure 2. This information includes the number of magnet, charter, special education, and private schools inside an attendance boundary.

BLOCK RECTIFICATION FOR CUSTOM CENSUS TABULATIONS

The primary set of geography SABINS supplies to the public are "block-rectified" attendance boundaries. In the SABINS database, attendance boundaries are aggregates of census blocks. Most attendance boundaries are closely aligned with the TIGER/Line files and, because these line features comprise census blocks, most attendance boundaries are, in fact, meant to entirely contain census blocks. Still, some school districts delineate some of their attendance boundaries so that a portion of them serves children on both sides of a street. Such attendance boundaries legitimately and intentionally split census blocks. Still, the SABINS database assigns an entire census block to an attendance boundary regardless of whether it is split by that attendance boundary. Thus,

---

9   When data are received as a union (and thus the topology is enforced) the tolerance level is set at zero. The most prominent examples are GIS data we received from the three states. These states stored their data as a union and all three states digitized their data using local cadastral data. A tolerance level of zero preserved the original positional accuracy of these data.

10   At this stage, it is also possible to extract the topologically corrected "elementary," "middle," and "high" school layers (and their corresponding NCESSSH school identification codes) from the union. This can be done by dissolving the union on the identification code assigned to each of the input layers. This allows users to obtain boundaries that cover grade-spans (e.g., the "elementary" grade spans as defined by a school district) rather than creating grade-specific boundaries.

Percent Within	Kindergarten	Grade 7	Grade 12
99 percent within	96.4	98.2	98.2
95 percent within	98.1	99.3	99.4
90 percent within	98.5	99.4	99.5
Number of Attendance Boundaries	84.0	33.0	23.0
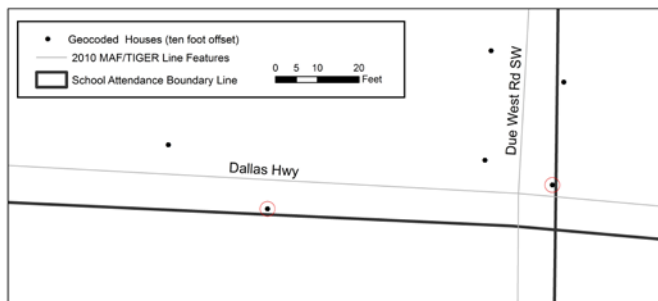
Figure 4.

block-rectified attendance boundaries are not precisely the same as those delineated or described by a local school district. Upon request, SABINS distributes the union file that has not been block rectified.

The SABINS project uses a straightforward block-rectification technique. The first step is to create a point file that represents the geographic center (or centroid) of all U.S. census blocks. The Census Bureau's block file contains the centroid coordinates for each block and these are used to create a point layer. The block point file also contains school district code(s). Next, a list of all unique school districts in the unioned polygon data is created. Each value from that list is used in a selection query that is applied to both the census blocks and the unioned polygons. For example, the first school district in the list is identified and used to select the census blocks that lie within the selected district and polygons from the unioned dataset that are also within the selected district. Then, for the selected district, ArcGIS 10.0's "near tool" is used to associate a block with an intersection in the unioned dataset. If a block centroid falls within a unioned polygon, that polygon is considered the "nearest." If a block centroid falls outside of a unioned polygon, it is assigned the feature ID of the closest unioned polygon. The near process ensures that all census blocks in a given school district (as defined by the Census Bureau) are assigned to a unioned polygon. The "selection query" and the "near process" are executed iteratively—once for each school district in the unioned dataset.

Once the block points have been assigned attendance boundary identification codes, the block points are rejoined to the original block polygons from which these points were generated. After the block polygon file is associated with the identification codes of an attendance boundary, the block polygons are dissolved into "block-rectified" attendance boundaries. As discussed below, because most blocks have more than 90 percent of their area within

an attendance boundary, the block-rectification process essentially conflates attendance boundaries with 2010 TIGER/Lines.

There are two reasons why the SABINS database assigns entire blocks to attendance boundaries. First, block-rectified boundaries can be used to create custom tabulations of census population counts that can be released to the public. The Census Bureau's disclosure policies include a stipulation that census blocks are entirely nested within the custom geography before a custom tabulation can be released publicly.

Second, block-rectified attendance boundaries conflate the line work of attendance boundaries with census blocks. Many school districts that supply electronic GIS files to the SABINS project use local cadastral data to delineate their attendance boundaries—as is the case for the districts in Delaware. An attendance boundary may follow a line segment such as a road, but this road, as delineated using local source data, is not aligned with TIGER/Line files. Thus, using the TIGER lines to address-match will locate some address points within the wrong (pre-block-rectified) attendance boundary (see Figure 4). This is shown by the addresses circled in red—which are assigned to the wrong boundary. After block rectification, a census block along the periphery of an attendance boundary will share geometry with the TIGER/Lines. This makes geocoding with TIGER/Line features more manageable. Of course, some addresses will not be accurately assigned to school boundaries after boundaries are rectified to blocks. These errors will occur when a school boundary intentionally splits a block.

DEGREE OF BLOCK NESTING

Table 4 shows the percentage of 2010 census blocks in Delaware that have varying proportions of their areas within an attendance boundary. Only blocks with at least one person living in them are considered. There are 15,933 populated blocks in Delaware. Of these blocks, 96.4 percent have at least 99 percent of their area within a kindergarten attendance boundary. This same figure is 98.1 percent for grades 7 and 12. If the proportion of a block's area within an attendance boundary is increased to 90 percent, then the percentage of blocks that are nested within kindergarten attendance boundaries is 98.5 percent; this figure is 99.5 percent for grades 7 and 12. The threshold of 90 percent is somewhat arbitrary. Still, if an attendance boundary contains 90 percent of a block's area, the imperfect nesting almost always results from discrepancies in line work between TIGER/Line features and locally defined features. This still means that roughly 2 percent

Table 5. Correlation coefficients between allocated and known percentages of various racial groups in attendance boundaries, Delaware, 2009–2010

Racial Group	Kindergarten	Grade 7	Grade 12
Non-Hispanic Black	.997	.999	.998
Hispanic	.993	.998	.998
Non-Hispanic White	.997	.999	.998
N Catchment Areas	84.0	33.0	24.0

of blocks are legitimately split by attendance boundaries. In cases in which a block is intended to be split by an attendance boundary, the entire block still is assigned to a single attendance boundary and, although this is less than desirable, it is necessary for obtaining custom tabulations from the U.S. Census Bureau.

It is important to know how much assigning entire blocks to attendance boundaries impacts population estimates within attendance boundaries. As in a previous analysis, grades kindergarten, 7, and 12 in Delaware are used to explore this topic. This analysis compares the percent difference of non-Hispanic white, non-Hispanic black, and Hispanic people in: (1) block-rectified attendance boundaries with (2) a modified areal weighting approach. The modified areal weighting approach assigns all of a block's population to an attendance boundary if the attendance boundary contains more than 90 percent of the block. If an attendance boundary contains between 10 and 90 percent of a block's area, the attendance boundary is assigned population totals in proportion to the area of the block that it contains. For example, if an attendance boundary contains 85 percent of a block, then it is assigned 85 percent of its population. Block-rectified boundaries consist of the entire population of whole census blocks—where a census block is assigned to an attendance boundary if its centroid lies within it.

The difference between the percent of non-Hispanic white people in boundaries that are block rectified and areal weighted is less than one percentage point for 93 percent of kindergarten boundaries. Slightly fewer than 5 percent of boundaries differ by 1 or 2 percentage points while about 2 percent of the boundaries differ by 2 to 4 percentage points. For 7 and 12 grade attendance boundaries, 100 percent of attendance boundaries have less than a 1 percentage point difference in the percent of non-Hispanic white people in block-rectified and modified areal-weighted attendance boundaries.

INTEGRATING DEMOGRAPHIC DATA WITH ATTENDANCE BOUNDARIES

The SABINS database provides users with demographic estimates describing the characteristics of persons, families, households, and housing units within block-rectified attendance boundaries. Data that describe these characteristics originate from two sources. The first source is the decennial census that summarizes complete-count population data at the census block level. The second is the American Community Survey (ACS). While the ACS provides a great deal of detailed information about the U.S. population, the challenge is to summarize these data to block-rectified attendance boundaries. The SABINS project uses a straightforward interpolation procedure to estimate sociodemographic characteristics in attendance boundaries based on ACS block group characteristics. The procedure first derives an ACS-based block group-level proportion of persons, households, housing units, or families with a given social characteristic. That proportion is multiplied by the appropriate 2010 census-based 100 percent count block variable for all census blocks falling inside the block group. Essentially, the procedure assumes that, for a given subgroup of the population, the proportion of the whole population in that subgroup is the same for all blocks in a block group. The SABINS project then re-aggregates the block-level interpolated values to attendance boundaries.

A simple hypothetical example describes the process of estimating the number of households with income below the poverty line for attendance boundaries. A block group contains three census blocks within it. In the 2010 Census, the first block had 100 households, the second 200, and the third 220. The 2007–2011 ACS estimates that the block group contains 500 households, with 100 households having an income below the poverty line. Thus, the 2007–2011 proportion of households with an income below the poverty line is 0.2. The SABINS project multiplies that proportion by the 2010 household counts in each block, yielding 20 households for the first block, 40 for the second, and 44 for the third. The block-level estimates of households with an income below the poverty line then are summed to attendance boundaries.

This allocation procedure, like all interpolation methods, introduces error into the estimates. The allocation procedure used in the SABINS project violates two assumptions: (1) the proportions of the population in a subgroup differ among the block groups within a particular block group; (2) the temporal mismatch between the ACS data and the decennial census yields different proportions because people move, households form and dissolve, and new housing units are built.

To determine how much error is introduced in the allocation procedure, we correlated the *actual* racial characteristics of people in attendance boundaries with values *interpolated* from block groups. Because 2010 census data provide counts of people by race at the block level, it is possible to generate actual counts of people by race for each attendance boundary. This actual count is correlated with estimates produced by interpolating block-group

data. Specifically, correlations between known and interpolated values were created for the percent of non-Hispanic white, non-Hispanic black, and Hispanic people in kindergarten, 7th grade, and 12th grade attendance boundaries in Delaware. Results are shown in Table 5. Findings indicate that correlation coefficients between the actual and interpolated values are at least .993 for all racial comparisons—and at least .997 for 8 of the 9 racial comparisons.

FROM GRADE-SPECIFIC ATTENDANCE BOUNDARIES TO GRADE-SPECIFIC SCHOOLS

Although schools provide services to attendance boundaries, schools and attendance boundaries do not have a one-to-one relationship. In some cases, boundaries are served by multiple schools. If a user wants to determine the population characteristics of the people living in the boundaries served by each school (for a single grade level), it is necessary to sum populations living in each school's attendance boundary and then divide these sums by the number of schools that serve the boundary.

This procedure is relatively straightforward. Users join counts of people within attendance boundaries to the schools that supply services to those boundaries. Once every boundary is joined to the school (or schools) that provide services to it, a value is generated that counts the number of schools that supply services to an attendance boundary. If one school provides services to an area (which is the most typical scenario), then the "school count" value will be one; if two schools supply services to an area, the count will be two, and so on. The "school count" value then is divided into the population counts of each school boundary and the "weighted counts" are summed to the schools that provide services to each boundary. This procedure preserves the original population counts of the entire school district while still providing the ability to produce meaningful statistics at the school level (e.g., the percent of children who are low income).

FROM GRADE-SPECIFIC SCHOOLS TO ENTIRE SCHOOLS

Many users will want to estimate the population characteristics of people who live within a school's catchment areas—not simply for a single grade but for an entire school irrespective of the grade span(s) it serves. To illustrate this process, assume that the goal is to generate the *proportion*[11] of black and white people who live within the attendance boundaries served by schools and to generate these estimates for all schools irrespective of their grade ranges (and ir-

respective of the fact that a school can serve different boundaries at different grade levels). The SABINS database consists of 13 grade-specific polygon layers spanning grades K to 12. Creating school counts from these 13 layers entails the following: First, the number of black and white people in each grade-specific set of boundaries is divided by 13. This essentially allocates 1/13th of the population to each of the 13 grades kindergarten to 12. The second step is to join the weighted counts of black and white people to the schools that provide services to those boundaries. Third, the (weighted) number of black and white people who live within each grade-specific attendance boundary is divided by the number of schools that provide services to it. Fourth, the 13 grade-specific, school-based data files are appended together (i.e., stacked on top of one another). For example, if a school serves five grades, the data for the five grades will be repeated in the database. The final step is to sum the number of black and white people across the 13 "stacked" files to each school. This is accomplished by aggregating (or collapsing) on school identification codes (using the NCESSCH field). This last step reproduces the original population counts in a school district, but the final result allocates data to whole schools and not simply to grade-specific boundaries. The result is an estimate of the characteristics of people who live within the attendance boundaries served by every school—irrespective of the grade spans and attendance boundary combinations that those schools serve. These steps can be completed in most statistical software.[12]

CONCLUSIONS AND FUTURE WORK

The School Attendance Boundary Information System is a spatial data infrastructure project that has, for the first time, collected, processed, harmonized, and disseminated K-12 educational geography on a massive scale. Spatial and tabular data can be downloaded from www.sabinsdata.org. The SABINS Web site contains tutorials describing how to use the data described in this paper.

The first major achievement of SABINS is to demonstrate the feasibility of collecting and digitizing attendance boundaries from school districts throughout the United States. The second accomplishment of the SABINS project is modeling some of the seemingly intractable relationships among attendance boundary geographies and the schools that supply services to those geographies. While it is reasonable for some school districts to treat schools and attendance boundaries as the same entities, such a system is useless at larger scales. Designing a robust database management system consisting of normalized, relational tables linking schools with boundaries allows for easy management, update and analysis of the database, and, more importantly, it

---

[11] The procedure described in this paragraph will work for descriptive statistics such as means and proportions—but not for counts of people. To determine the number of people in an area that a school serves, users have two basic options. The first is to use grade-specific boundaries. The second option is to work with data that describes the characteristics of students who are enrolled in specific grades and who live in specific attendance boundaries.

[12] Users can also create school-specific boundaries—that is, one boundary for every school. This can be accomplished using the "make-query" table function in ArcGIS 10.1. Since boundaries overlap, this will result in "stacked" polygons and this is not ideal but perhaps useful in limited circumstances. It also assumes that each school serves only one boundary—which is not true.

makes attendance boundary data usable by academic researchers and public policy analysts. This model serves as a template that states can adopt, modify, and incorporate into their own enterprise GIS systems.

Third, SABINS also uses straightforward spatial interpolation techniques that estimate the sociodemographic characteristics of people and households located within school boundaries. Although allocating demographic information from census geography to attendance boundaries is reasonably accurate, it is important to remember that attendance boundaries are "permeable" because some students are enrolled in private, charter, and magnet schools that draw children from traditional, neighborhood schools. While the interpolation techniques used in the SABINS project result in reasonable estimates of the characteristics of people who live within school boundaries, these estimates are imperfect reflections of who is enrolled in a school. Nevertheless, this will be partially addressed with the release of custom tabulations from the U.S. Census Bureau that consist of public school children only—which is a closer representation of who is enrolled in the school that serves a given boundary. Taken as a whole, SABINS provides researchers, policy makers, and local administrators with a rich new spatial and tabular data source that serves the diverse needs of a wide constituency.

Finally, the SABINS project has been institutionalized and currently is under the auspices of the U.S. Department of Education and the U.S. Census Bureau. These agencies will collect attendance boundaries for the entire country for the 2013–2014 and 2015–2016 school years. Complete coverage of the United States will be supported by a Web-based, digitizing service that will allow school districts to digitize their boundaries remotely. The remote-digitizing system has several advantages. Local districts can save money for they do not have to buy software or pay consultants to digitize and display their boundaries; it allows the project to collect data on an even larger scale and in a standard format; and it improves the accuracy of the data because the system can provide tutorials that teach best practices in using GIS.

Acknowledgments

About the Authors

Salvatore Saporito is an associate professor of sociology at the College of William and Mary. His research investigates the causes and consequences of racial and economic segregation in schools and attendance boundary information lies at the heart of much of his work. His latest research investigates the impact of attendance zone gerrymandering on racial and economic segregation. He has a B.A. in sociology from Glassboro State College and a Ph.D. in sociology from Temple University.

Corresponding Address:
Department of Sociology
The College of William and Mary
Williamsburg, VA 23187-8795
E-mail: sjsapo@wm.edu

David Van Riper is the Director of the Spatial Analysis Core at the Minnesota Population Center and has worked extensively on creating and disseminating large GIS data infrastructure such as the National Historical Geographic Information Systems (http://www.nhgis.org). He has completed a B.A. in geography in 1999 from the University of Wisconsin-Madison and holds a M.A. in geography from the University of Minnesota.

Minnesota Population Center
50 Willey Hall
225 – 19th Avenue South
Minneapolis, MN 55455
E-mail: vanriper@umn.edu

Ashwini Wakchaure is a GIS programmer on the SABINS project. She holds a Ph.D. from the Department of Urban and Regional Planning at the University of Florida.

Spatial Sciences Institute
University of Southern California
Los Angeles, CA 90089-0374
E-mail: ashwini.wakchaure@usc.edu

References

Alexander, K., D. Entwisle, and L. Olson. 1997. Children, schools and inequality. Boulder, CO: Westview Press.

Black, S. 1999. Do better schools matter? Parental valuation of elementary education. Quarterly Journal of Economics 114: 577-99.

Brunner, E., J. Murdoch, and M. Thayer. 2002. School finance reform and housing values: Evidence from Los Angeles. Public finance and management 2: 535-65.

Brunner, E., J. Sonstelie, and M. Thayer. 2001. Capitalization and the voucher. Journal of Urban Economics. 50: 517-36.

Card, D., and A. Kreuger. 1992. Does school quality matter? Returns to education and the characteristics of public schools in the United States. Journal of Political Economy 100: 1-40.

Chen, C. 2011. Numbers and types of public elementary and secondary schools from the Common Core of Data: "School year 2009–10 (NCES 2011-345)." U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved June 15, 2011, from http://nces.ed.gov/pubsearch.

Diez Roux, A. 2000. Investigating neighborhood and area effects on health. American Public Health Association 91: 1,783-89.

Downes, T., and J. Zabel. 2002. The impact of school characteristics on house price: Chicago 1987–1991. Journal of Urban Economics 52: 1-25.

Edwards, V., and M. Ehrenthal. 2008. Making individual school enrollment projections using "micro-geographies." Presentation at the Population Association of American Meetings, New Orleans. April 17 to 19.

Elliott, P., and D. Wartenberg, 2004. Spatial epidemiology: Current approaches and future challenges. Environmental Health Perspectives 112: 998-1,006.

Frankenberg, E., C. Lee, and G. Orfield. 2003. A multiracial society with segregated schools: Are we losing the dream? Cambridge, MA: Harvard University Civil Rights Project.

Heckman, L., and H. Taylor. 1969. School rezoning to achieve racial balance: A linear programming approach. Socio-Economic Planning Sciences 3: 127-33.

Huang, R., and D. Hawley. 2009. A data model and Internet GIS framework for safe routes to school. Journal of the Urban and Regional Information Systems Association 21: 21-31.

Ionnides, Y. 2004. Neighborhood income distribution. Journal of Urban Economics 56: 435-57.

Krieger, N. 2006. A century of census tracts: Health and the body politic (1906–2006). Journal of Urban Health 83: 355-61.

Krieger, N., P. Waterman, J. Chen, M. Soobader, S. V. Subramanian, and R. Carson. 2002. ZIP code caveat: Bias due to spatiotemporal mismatches between ZIP codes and U.S. census-defined geographic areas—the public health disparities geocoding project. American Journal of Public Health 92: 1,100-2.

Lemberga, D., and R. Church. 2000. The school boundary stability problem over time. Socio-Economic Planning Sciences 34: 159-76.

Logan, J., and D. Oakley. 2004. The continuing legacy of the Brown decision: Court action and school segregation, 1960–2000. Albany, NY: Lewis Mumford Center for Comparative Urban and Regional Research.

National Academy of Sciences. 2009. Estimating children eligible for school nutrition programs using the American Community Survey. Current projects system. Retrieved from http://www8.nationalacademies.org/cp/projectview.aspx?key=49102.

New Mexico Statute Section 14-2-6(E), NMSA 1978.

Owens, A. 2010. Neighborhoods and schools as competing and reinforcing contexts for educational attainment. Sociology of Education 83: 287-311.

Panel on Estimating Children Eligible for School Nutrition Programs Using the American Community Survey; Committee on National Statistics, Division of Behavioral and Social Sciences and Education; National Research Council. 2012. In Shirm, A., and N. Kirkendall, Eds., Using American community survey data to expand access to the school meals programs. Washington, DC: The National Academies Press.

Reardon, S., and J. Yun. 2001. Suburban racial change and suburban school segregation, 1987–1995. Sociology of Education 74: 79-101.

Sampson, R., J. Morenoff, and T. Gannon-Rowle. 2002. Assessing "neighborhood effects": Social processes and new directions in research. Annual Review of Sociology 28: 433-78.

Saporito, S. 2009. School choice in black and white: Private school enrollment among racial groups, 1990–2000. Peabody Journal of Education 84: 172-90.

Shai, D. 2006. Income, housing, and fire injuries: A census tract analysis. Public Health Reports 121: 149-54.

U.S. Census Bureau. 2011. 2010 Topologically integrated geographically encoded referencing system line shapefiles (TIGER/Line Shapefiles). (Machine readable data files.) Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. 2011. 2006–2010 American community survey (ACS). (Machine readable data files.) Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. 2011. 2006–2010 American community survey (ACS) 5-year summary file technical documentation. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. 2011. 2010 Census summary file 1 (SF1.). (Machine-readable data files.) Washington, DC: U.S Census Bureau.

U.S. Department of Education. 2011. Common core of data. (Machine-readable database.) Washington, DC: U.S. Department of Education.

U.S. Department of Education. 2008. Private school universe survey. (Machine-readable database.) Washington, DC: U.S. Department of Education.

Weimer, D., and M. Wolkoff. 2001. School performance and housing values: Using noncontiguous district and incorporation boundaries to identify school effects. National Tax Journal 54: 231-53.

Winkleby, M. A., and C. Cubbin. 2003. Influence of individual and neighborhood socioeconomic status on mortality among black, Mexican-American, and white women and men in the United States. Journal of Epidemiology and Community Health 57: 444-52.

Xue, J., T. McCurdy, J. Burke, B. Bhaduri, C. Liu, J. Nutaro, and L. Patterson. 2009. Analysis of school commuting data for exposure modeling purposes. Journal of Exposure Science and Environmental Epidemiology 20: 69-78.

Improving Cadastre: Development of a Workflow Prototype Utilizing ESRI's Parcel Fabric

Linda M. Foster and Justine I. Blanford

Abstract: *The Rapid City/Pennington County, South Dakota, GIS Division has continuously maintained a parcel dataset that was originally created in 1989. Advances in technology and the desire to expand the use of land-based information, requiring highly accurate data, highlighted the need to improve the cadastre. Technical obstacles, such as incorporating and maintaining survey information as well as easily updating related layers, previously hindered this effort. The Environmental Systems Research Institute (ESRI) parcel fabric data model purports to have resolved many of these issues and was the focus of this project. A workflow was developed and an accuracy assessment performed during this study to test these claims. The project successfully addressed the following key factors: (1) conversion of the existing data, (2) maintaining and improving the integrity of the cadastre data over time, and (3) integrating these data with related data layers. Although the workflow was developed for Rapid City, it is scalable and it can be applied by any organization that maintains land records.*

INTRODUCTION

Efficient and effective management of limited resources, such as land, is becoming more and more important as the United States continues to grow and development densities compound. Rapid City, South Dakota, not unlike many other communities, uses geographic information systems (GIS) to manage its land records (cadastre[1]) and other spatial information. For example, its parcels dataset is used to maintain ownership and tax information, record zoning and other planning designations, track annexations, maintain corporate boundaries, and develop future land-use plans. To date, the cadastre parcels are a representation containing accurate attribute information about the land such as area, ownership, and tax value. Historically, there was no need for accurately surveyed spatial data because it was developed primarily for taxation purposes and little if any other relevant spatial data existed. However, in recent years, additional datasets such as high-quality aerial imagery and sanitary sewer infrastructure have been developed with high positional accuracy. These layers are constantly under consideration by engineering and planning staff, and when plotted with base layers, such as the parcels, disparities in accuracy between the datasets become apparent, thus highlighting the need to improve the accuracy of the cadastre layer. Two layers in particular are driving the city's interest in improving its parcel base: zoning and future land use. Having these layers available and up-to-date would increase staff efficiency when reviewing development submittals, improve customer service by having the data accessible to the public, and help expedite planning and engineering studies.

History of Cadastral Dataset of Rapid City

The original cadastral dataset for Rapid City was developed in 1989 (see Figure 1 for overview) from plats at three scales and adjusted to U.S. Geological Survey (USGS) 7.5 minute quadrangle section corners resulting in some errors. From 1989 through 2000, parcels were added by digitizing and using coordinate geometry (COGO) input methods. Rectified but not ortho-corrected aerial images also were used to help align the property lines. As new imagery was acquired, many lines had to be adjusted, especially in areas of high relief (Rapid City GIS Division 2009). In 2000, the parcels were converted to the Environmental Systems Research Institute (ESRI) ArcInfo Coverage format, and again some errors were introduced. According to the GIS Division staff, there was reasonably good conversion of the data in the eastern half of the county but less so in the western half (Rapid City GIS Division 2009). Not only were errors introduced during the conversion, but sometime after the project was finished, it also was discovered that the conversion vendor incorrectly moved section lines to match the digital line graph (DLG) section lines, rather than moving the parcels to the correct section. In addition, water boundaries were erroneously incorporated to represent parcel boundaries. In 2003, the ESRI ArcInfo parcel coverages were converted into one contiguous countywide ArcSDE feature class. Maintenance of the parcels has continued using ESRI's ArcMap desktop software by COGO input and other editing techniques.

From the original development of the parcels dataset through the conversions discussed above, errors have been introduced and continue to be propagated. Even the current methods used for
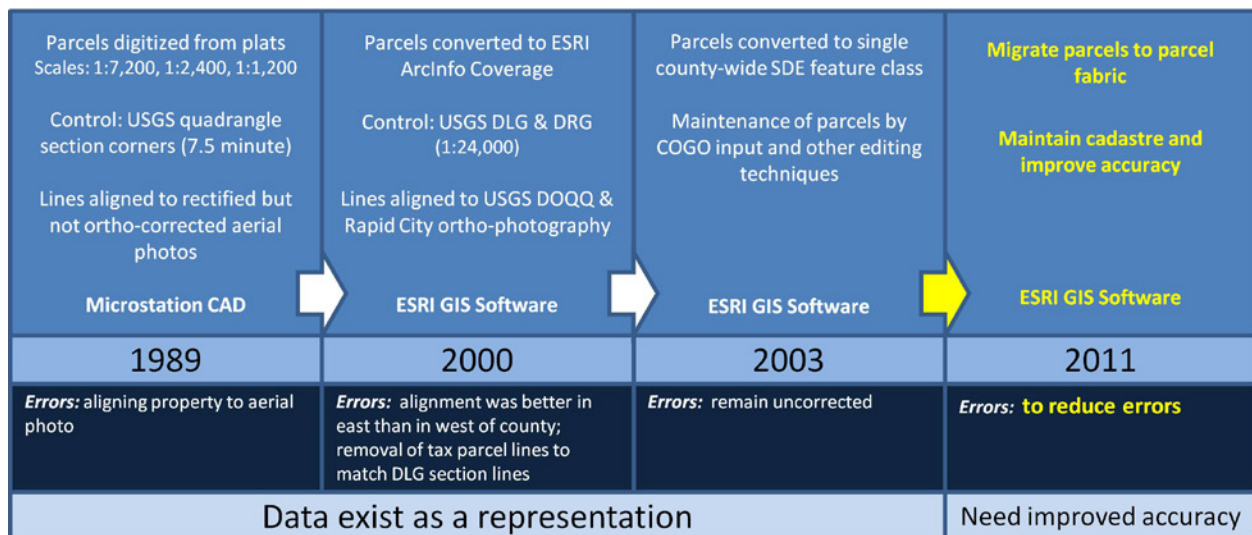
Figure 1. History of the creation and evolution of the dataset of the Rapid City parcels
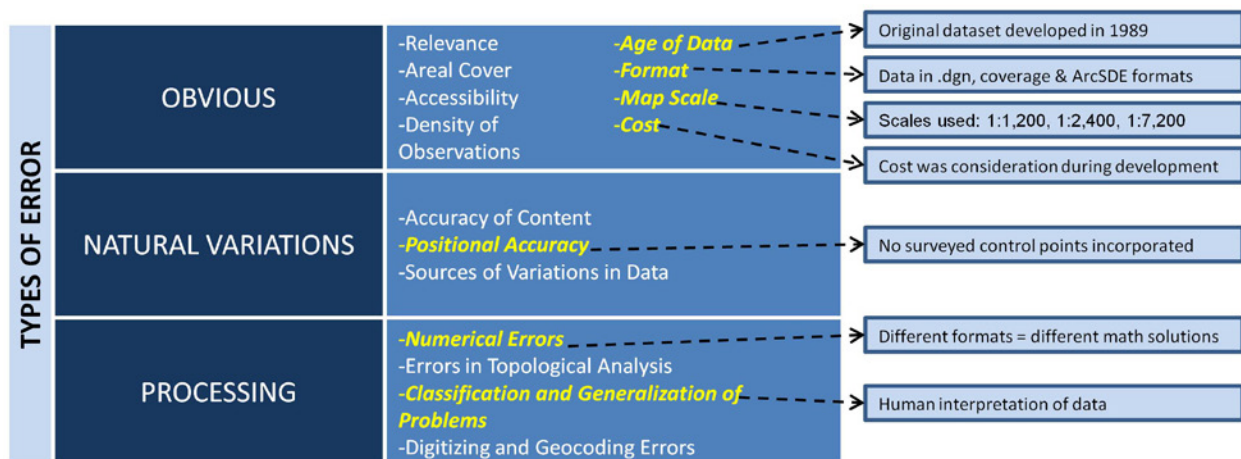
Figure 2. Sources of error commonly found in geospatial datasets (Foote and Huebner 1995) and how they apply to the dataset in this study

updating and maintaining the data introduce, if not maintain, error in the dataset. For example, when an area is newly subdivided, the surveyor of record's platted information is reproduced digitally using software with coordinate geometry (COGO) input capabilities. Data integrity then is often compromised so that the shape(s) can fit into the area available in the parcel layer instead of being truly represented.

Historically, one of the main factors limiting spatial accuracy in GIS systems was the capacity of hardware and software and their inability to handle geodetic coordinate systems effectively. However, as both of these have improved, this no longer is a limitation. The wide availability and substantial improvements in spatial data quality provided by global positioning systems (GPS), aerial photography, and other data-collection technologies have found spatial management and improved accuracy of cadastral databases struggling to keep pace (Harper 2006).

Although errors are naturally inherent in geospatial data, data

collected by observation tends to suffer from imperfect quality more than other types of data as a result of subjective interpretation rather than precise measurement (Goodchild 1992). Foote and Huebner (1995) highlight three types of errors associated with geospatial data that are summarized in Figure 2. Several of these are present in Rapid City's cadastre (highlighted in yellow, Figure 2) and include obvious errors (age of data), natural variations (positional accuracy), and errors caused by processing of the data (numerical errors and geocoding and digitizing errors).

Once the sources of errors have been identified, making changes to the parcels dataset, whether to accommodate the dynamic nature of land configuration or make adjustments to improve accuracy, currently poses a problem. Handling other land-dependent layers, such as zoning, future land use, street centerlines, corporate limits, annexation boundaries, and utility features, becomes very resource-intensive if all changes being made to the land base are to be reflected in the associated layers.
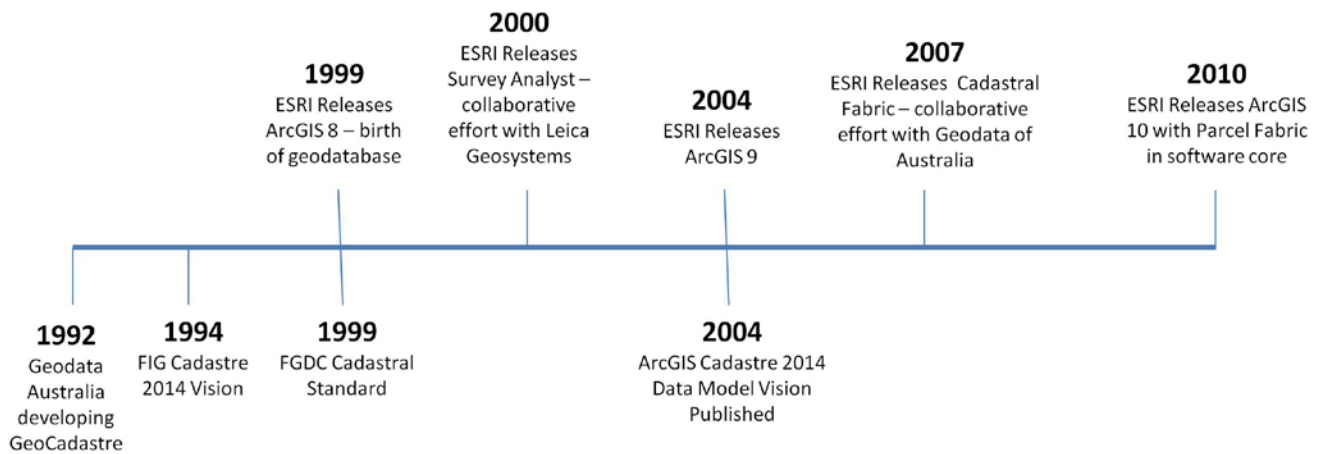
Figure 3. Generalized timeline of the introduction of cadastral standards, collaborations, and products

Historically, these changes have not been consistently maintained in the associated layers, producing a less than visually appealing result when the parcels are overlaid and troubling results when some spatial data analyses are performed.

Rapid City has several common cadastral objectives that include the development of cadastral layers with higher spatial accuracy, applying cadastral adjustments to associated layers, increasing accuracy over time by continuous updating and maintenance, and storing legacy data within the cadastre fabrics (Bhowmick et al. 2008). ESRI's parcel fabric data model appears to meet these objectives.

ESRI's cadastral solutions, including the parcel fabric data model, have been in development for quite some time and are the result of multiple collaborations. The data model was crafted to consider the objectives of the Cadastre 2014 Vision set forth by the International Federation of Surveyors (FIG) Commission 7 group and the Federal Geographic Data Committee (FGDC) Cadastral Data Content Standard for the National Spatial Data Infrastructure (Kaufmann and Steudler 1998, ESRI and Kaufmann 2004). Figure 3 is a generalized timeline of the introduction of cadastral standards, collaborations, and products leading up to the integration of the parcel fabric data model in ESRI's current software core.

In 2010, ESRI renamed cadastral fabric to parcel fabric, and changed the related tools and editing technology from an extension product to a part of the core ESRI software. The parcel fabric technology, which is the focus of this project, is the result of more than two decades of research and development by ESRI and its partner Geodata of Australia (Geodata 2006). Careful consideration given to national and international standards, decades of development, and the successful implementation of the GeoCadastre process in other countries (e.g., Australia, New Zealand (GeoData 2006); Vietnam (Huong 2010); United States (Florida: Capobianco and Mann 2009; Denver: Genzer and Tessar, 2011); see Konecny 2011 for overview of variation in land-management

systems in diverse geographic regions) signifies that a potentially stable, comprehensive solution has been developed. ESRI committing to this model and incorporating this package into the standard GIS software provides further confidence that this is a model/framework that was developed with longevity in mind.

In essence, the model fits parcels into their appropriate locations in the fabric based on points the parcel has in common with the fabric. Once this has been accomplished, the fabric then can be associated with other layers, reducing discrepancies and mismatching of boundaries. Not only does the parcel fabric resolve the aforementioned issues, but it also allows for preservation of historical data (i.e., maintains records of previous transactions enabling the user to review the state of a chosen area over time) and the maintenance of data in multiple projections.

The ability to import existing data and improve it over time is very important to the city of Rapid City from a feasibility standpoint. Some have alluded that existing datasets should not be salvaged and continue to be improved upon, but rather the fabric should be built from scratch to ensure its integrity (Harper and Lee 2008). For a GIS Division with a full-time staff of three supporting both county and city GIS activities, it is simply not reasonable to use this approach. Rapid City's parcels dataset, which has been used largely as a representation, has served its original purpose. However, with the advancement of spatial data technologies and an increasing integration of digital data systems into daily workflows, the city and its stakeholders have expressed a desire to improve the accuracy of the parcels dataset and related base data layers. The remainder of this paper will outline and evaluate a workflow for preparing and importing existing data into the parcel fabric, adjusting the parcels to control points, performing an accuracy assessment of the adjustment, and applying the adjustments to an associated layer.
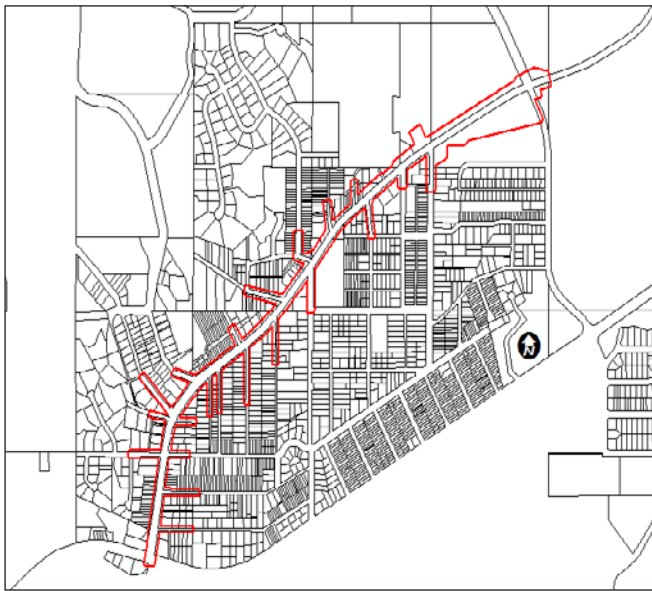
Figure 4. Map illustrating the cadastral data layer that will be used for this study. The Canyon Lake Drive neighborhood area contains approximately 675 parcels. The red line delineates the street reconstruction project zone and Canyon Lake Drive.

METHODOLOGY

Data

The data used for this study was a small portion of parcels from Rapid City's existing cadastre. The area was chosen because a major arterial street reconstruction project was recently completed in the area (see Figure 4), providing an ideal comparison dataset for use in this study. During the design phase of this street project, an accurate property layer had to be assembled so properties impacted by construction activities could be identified. Detailed property information also was necessary for developing construction easement documents and acquiring necessary rights-of-way. To develop the property layer, property corners in the project area were located and recorded using a mix of GPS and conventional surveying methods. Plats, easements, deeds, and other existing property documentation were retrieved from the county courthouse. A cadastral layer for the project area then was constructed in AutoCAD Civil 3D 2011, using the plats and surveyed property corner information. For this parcel fabric study, the surveyed property corners provided geodetic coordinates for import into the fabric to adjust the existing parcels to. And having the independently created cadastral layer provided an opportunity for a comparison to see how well the parcel fabric adjusted the parcels in the test area.

WORKFLOW

Five steps identifying the workflow necessary to test and implement the parcel fabric for Rapid City have been identified and are summarized in Table 1; they will be discussed in more detail in the following sections.

STEP 1	Building Framework
STEP 2	Preparing and Loading Data
STEP 3	Parcel Adjustment
STEP 4	Compare adjusted parcels to AutoCAD cadastre
STEP 5	Adjusting an associated layer

Table 1. Five-step workflow developed during this study

1.	LINES	Must Not Self-Overlap
2.	LINES	Must Not Self-Intersect
3.	LINES	Must Not Intersect or Touch Interior
4.	LINES	Must Be Covered by Boundary Of
5.	POLYGONS	Boundary Must be Covered By
6.	LINES	Must be Single Part

Table 2. Topology rules required at a minimum by the parcel loader to load data to the fabric

### Step1: Building the Data Migration Framework

The first step in the workflow development of this project was reviewing existing documentation to identify the necessary steps required to prepare the data for loading into the parcel fabric. This included reviewing ESRI documentation and other available literature (which is limited for this is still a relatively new component) as well as conversing with ESRI personnel. The workflow step developed consisted of approximately 12 items. This includes technical and data-related tasks such as verifying software version, installing necessary components such as the Curves and Lines tool (ESRI 2010), creating workspaces, and verifying projection and coordinate system information. Feedback in the form of verbal communication was received from the end user and incorporated into the final workflow procedure.

Step 2: Preparing and Loading Data to the Parcel Fabric

The second step required preparing and loading data into the parcel fabric and documenting the steps involved. Using the rules required by the parcel fabric (see Table 2), a topology of the lots was created and successfully verified.

However, the parcel loader failed to load the lot lines, citing topology errors. Further investigation revealed that even though the data passed all the topology requirements, additional editing of the data was needed. This included using tools to planarize the lines (i.e., break at intersections (Figure 5A)) and split multisegment lines at inflection points (Figure 5B) (i.e., where a curve transitions into another curve, or at sharp bends, etc. (ESRI 2010)).

Tax parcels were loaded next. Assuming that the same approach the city had been using for deriving tax parcels (dissolv-
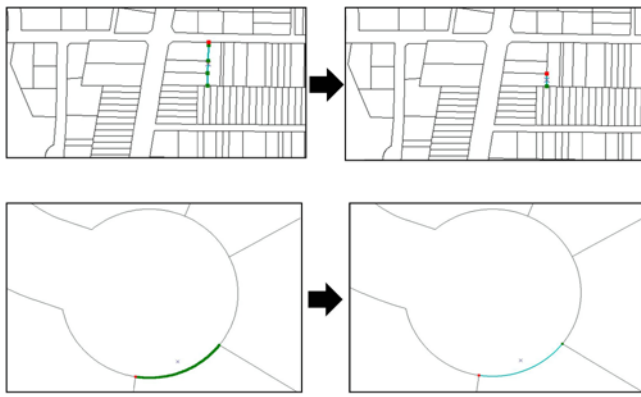
Figure 5. Examples of topology errors prior to (A) planarizing and (B) conversion of lines to two-point curves

Figure 6. Control points loaded into the fabric for the study area and associated with the appropriate parcel corners

Figure 7. A parcel in the parcel fabric resulting from the import of existing data (left) and an image of its corresponding plat (right)

ing by attribute) would be appropriate, tax parcels were derived from the lot lines that had been successfully loaded to the fabric. Again, a topology was created and verified and the tax parcels were loaded. The parcels were within the tolerances required by the parcel loader and loaded without error. However, on close visual inspection, the tax parcel lines were not coincident with the lot lines loaded previously. Apparently, the process of dissolving the features by attribute resulted in a slight amount of movement. The concern with the movement is mostly cosmetic in nature, for it did meet the tolerances required by the parcel fabric. However, Rapid City did choose to pursue another option that would result in coincident lines, as described next.

To address this issue, tax parcels were re-created from the lot lines. Two different approaches can be used to isolate the lot lines that need to be removed to derive the tax parcels. The first option is to simply order the layers in the *Table of Contents* of the project so that the tax parcels are on top of the lot lines and visually select all the lot lines that are not parcel boundaries and delete them. The other option, and one that will be more practical for Rapid City to use on the countywide dataset, is to use select by location with *Target layer(s) features are within (Clementini) the Source layer feature* option selected. This should result in most of the lot lines that are not tax parcel boundaries being selected, which can be deleted at the same time. However, if this method is used, it is important to check for lines that may have been erroneously removed. Two methods can be used here and include (1) by visually inspecting the layers and (2) comparing polygon counts with the original tax parcel layer.

Once the tax parcel lines and polygons were successfully loaded to the fabric (and checked), the control points were loaded to the fabric. Associations were made between the parcel corners and corresponding control points (see step 3 for more detail) (Figure 6).

Control points define accurate, surveyed x,y,z coordinates for physical features on the surface of the earth and in this study consisted of property corner monuments that had been located on the ground and coordinates recorded. While parcel dimensions accurately define parcel boundaries in relation to each other,
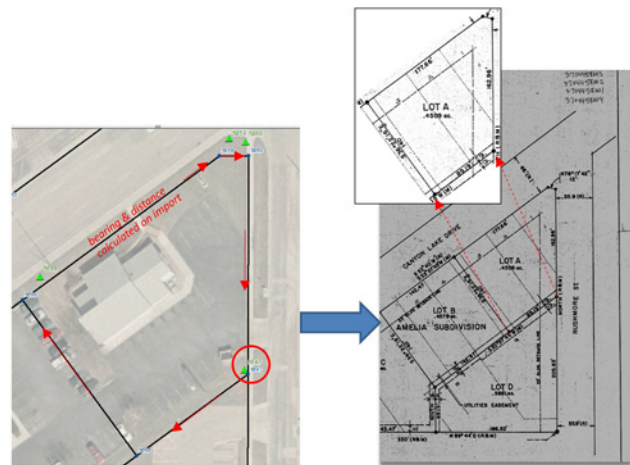
control points, when used in a least-squares adjustment, result in accurately defined spatial locations for parcel corner points (ESRI 2011), See step 3 for more details.

To summarize step 2: The existing parcels dataset consisted of parcel shapes without any coordinate geometry (COGO) attributes (i.e., bearing and distance of record) and did not necessarily truly represent the shape of the parcel (too many vertices and line segments making up the curves); it was processed and imported into the parcel fabric. The data-processing component of this workflow step consisted of breaking down these shapes into components that closely represent the platted shapes (see Figure 7) (i.e., two-point lines and parametric curves) and was accomplished through planarizing the lines and identifying the curves. The result was a fabric-ready set of lines, points, and polygons. The more closely each parcel represents its originally platted course, less editing and maintenance will likely be required once the data has been loaded to the parcel fabric (Denver GIS 2011).

Step 3: Adjusting Parcels to Control Points

As previously mentioned, the third step in the process is to use the least-squares adjustment built into the parcel fabric to adjust the existing parcels to surveyed control points. During this process, control point coordinate values are held fixed while the horizontal

Figure 8. An example of an area of parcels that is well balanced geometrically with evenly distributed control

and vertical coordinate system of the control points is transferred to the parcel fabric. In other words, control points are processed together with recorded dimensions to derive new, more accurate coordinates for parcel corners (ESRI 2011). Line dimensions (attributes representing the original survey) are not changed, but fabric point coordinates are updated and the geometric and spatial representation or the parcel line shape is updated. The result is an accurate coordinate-based cadastral system.

Least-squares adjustments are one of the most rigorous yet easy to apply without bias adjustments and are defined by Craig and Wahl (2003, p. 92) as being "based on the mathematical theory of probability and the condition that the sum of the squares of the errors times their respective weights is minimized." They also point out that one of the most important benefits of using the least-squares method of adjusting is that all types of survey measurements can be analyzed simultaneously.

In the parcel fabric, this adjustment is applied to a group of selected parcels and should be in an area that has a reasonably well-balanced geometric shape with redundant measurements (i.e., where multiple observations are made of the same point) and evenly distributed control (Figure 8).

Repeated observations validate a measurement network and a parcel fabric is a redundant measurement network. As pointed out by Craig and Wahl (2003, p.92), "Prudent surveyors check the magnitude of the error of their work by making redundant measurements."

Each parcel dimension and thus each parcel in the parcel fabric can have an associated accuracy. This is because parcel dimensions are derived from raw survey measurements, which have associated accuracies. By default, accuracy in the parcel fabric is defined by survey date because, in general, surveying equipment is more precise today than it was in the past, allowing for relatively greater accuracy in survey representations of parcel corners (ESRI 2011).

Accuracy assignments in the parcel fabric are important in the

least-squares adjustment because parcels with a higher accuracy assigned to them will have a higher weight in the adjustment and will adjust less than those parcels with lower accuracies. In other words, low-accuracy parcels will adjust around the more accurate parcels (ESRI 2011). ESRI uses seven accuracy levels with the highest level of accuracy given to the most recent surveyed data, mainly because of the ability for modern survey equipment and procedures to more accurately capture parcel data.

Data that were imported in previous steps of the workflow were automatically assigned an accuracy level of six, the lowest that can participate in an adjustment for the dimensions were calculated on import and not entered from a plat. If the data had been entered off a plat, then an accuracy level could have been assigned based on the date of the plat and would have ranged in accuracy between 5 ppm and 1,000 ppm.

Prior to running a least-squares adjustment, ESRI recommends checking the fit of control points. This calculates the transformation between the linked fabric point coordinates and the coordinates of the control points. The calculated parameters then are applied to the linked fabric point coordinates to compute temporary new values for the fabric point coordinates. The difference between the newly calculated fabric point values and the original control point values are reported as residuals for each active control point. Large residual values can indicate a problem in the data and should be investigated further. For instance, a large discrepancy (identified as being outside the range of the rest) may be the result of a poor control point, inaccuracy in the parcel data, or control points incorrectly matched to corresponding parcel points, and should be further investigated prior to applying the adjustment.

Perhaps one of the biggest drawbacks of the least-squares adjustment is that one wrong piece of information that goes undetected can greatly distort the results of the adjustment. This is because in the squaring process large residuals are dominant. A large measurement error that is ten times larger than the others will have the same effect on the sum of the squares as will 100 of the others (Craig and Wahl 2003). However, the dominant effect of squaring large residuals also enhances the ability to identify large errors that do not fit with the rest of the data and thus allows easier detection of mistakes that need to be removed or corrected. Therefore, it is imperative that the statistics be reviewed and suspect residual values addressed prior to committing an adjustment.

Step 4: Accuracy Assessment

The fourth step in the workflow is to perform an accuracy assessment of the adjusted data. In this case, an AutoCAD layer that was independently constructed from original plat documents and surveyed control points was used to make comparisons. Plotting the parcel fabric with the AutoCAD layer and visually inspecting how the two overlap was the first assessment of how well the adjustment performed. In areas where there is no independent work to check against, a visual inspection against aerial photography or other such imagery will provide some verification of the success of the adjustment. However, visual inspection of the data is

Table 3. Summary of ranking system used to evaluate success of least-squares adjustment

Rank	Percentage of Parcel Lines +/- 2.0 feet from Control Layer
1	100 – 90%
2	89 – 75%
3	74 – 50%
4	49 – 0%

a qualitative assessment; therefore, a more quantitative approach also was used. Twelve samples were taken with a range of seven to 44 parcels in size. Each sample was adjusted, using the least-squares method described above. Accuracy of fit was performed visually by viewing how well the parcel boundaries overlapped, analyzing the output statistics provided by the software, as well as ranking accuracy based on what percentage of the parcel lines were within +/- two feet of the control layer boundaries as described in Table 3. The percentage of parcel lines within +/- two feet of the control layer was determined by buffering the control layer lines and performing a spatial query.

Standards, such as the U.S. Geological Survey (USGS) National Map Accuracy Standards (NMAS)[2], were reviewed when considering an appropriate tolerance for evaluating the accuracy of parcel fabric adjustments. However, as pointed out by the International Association of Assessing Officers (IAAO), "The NMAS is most appropriate for paper maps which are only viewed at a printed scale,"[3] and also contends that "no one accuracy standard meets all needs" because of the differences between urban and rural environments (IAAO 2009, p. 10).

For this study, it was understood that the cadastral data will not be published at a static scale, the area under consideration is urban in nature, and Rapid City has obtained highly accurate utility location information that is helping drive the desire to improve related cadastral information. Knowing this, an accuracy range of +/- two feet was selected for evaluating parcel fabric adjustments for this project and was based partially on tolerances that were established by a parcel fabric project implemented by a utility company (Colorado Springs Utilities), whose cadastral products also are consumed by local government agencies (Moran et al. 2008). In summary, for the purpose of this study, accuracy was assessed using the ranking system summarized in Table 3.

An initial assessment of the 12 samples found that only 8.33 percent of the samples were fitting well (ranked 1 or 2) prior to any adjustments being applied. After the first adjustment was performed, this only increased to 25 percent of the samples. The reasons for this may be the result of a number of problems that include: incorrect shape of the parcel boundaries; inaccurate control points; inadequate control points; disproportionately distributed control points (i.e., larger number of control points on the perimeter of the sample and/or clustering of control points with large gaps between control points).

 (a) Incorrect shape of the parcel: If the shape of the parcel is incorrect, then the shape will need to be re-created using the original plat document and rejoined to the fabric. Obviously knowing this is difficult without a dataset for comparison, as has been done during this study. When the northing and

easting values are not converging to zero or stabilizing during the adjustment, this can indicate incorrect shape. Therefore, visual inspection against a control layer (as was done in this study) or aerial imagery can be used.

(b) Inaccurate control points: If a control point is problematic, it will need to be either corrected or deactivated and the adjustment reapplied. High or irregular residuals during the check fit indicate an inaccurate or incorrectly associated control point.

(c) Inadequate control points: There may be instances where there are few if any control points in an area that correction is desired. If there are none, obviously some will need to be acquired. If there are too few to perform the adjustment, or the points available are clustered, some additional points should be obtained to strengthen the adjustment.

(d) Distribution of control points: For cases where the distribution of control points is poor, additional control points will need to be added before applying an adjustment. Distribution of control points causing an adjustment to perform poorly can be identified by ruling out problems addressed in points (a) and (b) above. If neither control point accuracy nor shape appears to be an issue, then distribution of control points should be evaluated. If there are more control points around the outer edge of an adjustment area than inside, and the adjustment performed well around the outer boundary but not well internally, then it is reasonable to pursue adding some additional control points inside the adjustment area.

To identify what issues might be inhibiting the potential of the parcel fabric adjustment, each sample area was evaluated starting with the lowest ranking sample. Going through each of the steps listed above, a visual inspection was performed comparing the parcel shapes in the sample area to the control layer, assessing the reasonableness of the control point accuracy, and looking at the number and distribution of the control points. Notes were taken regarding what was observed and appropriate action taken (e.g., adding additional control points, inactivating bad control points, improving distribution of control points by adding more, etc.). In areas where the distribution of control points was obviously skewed (e.g., all control points located around the outer edges), an attempt was made to disperse the added points in as balanced a manner as possible.

Step 5: Adjusting an Associated Layer: Zoning

The fifth step in the workflow process is to apply the parcel adjustment to an associated layer. For this study, the zoning layer was chosen. If the desire is to adjust a parcel-based layer, such as zoning, it must be associated to the parcel layer being adjusted before the adjustments are performed. As such, the first step of this workflow was to verify that the zoning layer was associated with the parcels. After the parcels were adjusted, the adjustment vectors then were applied to the zoning layer, resulting in the zoning layer now

aligning with the parcels that were adjusted (sample 7, Table 4) and thus moved during this process (Figure 9). Seeing this happen successfully was a big victory because one of the biggest challenges the city faced adjusting parcels in the past was how to efficiently and accurately apply these improvements to related layers.

Figure 9. Zoning (shown in orange) prior to (left) and after (right) the parcel adjustment vectors were applied

RESULTS

The development of the workflow described in this paper has been an interactive and iterative process with Rapid City to ensure that the process can be executed successfully and applied to the remaining parcels for Rapid City and Pennington County (approximately 40,000 parcels). At the conclusion of each step, a written workflow process has been provided. All workflow steps have been tested by at least one staff member of the GIS Division for usability and feedback has been incorporated in the final workflow.

Based on the material presented in this paper and feedback received from the city of Rapid City GIS Division, the workflow that was developed as a result of this study has successfully met the objectives that were set forth for the project and included (1) developing a feasible workflow for converting existing data; (2) maintaining and improving the integrity of cadastre data over time; and (3) being able to integrate these data with related layers.

Accuracy of the parcels was greatly improved using a multi-step reiterative adjustment procedure as outlined in the methodology. During this study, two adjustments were required to reduce inaccuracies and are summarized in Table 4 below.

When the data was first loaded into the parcel fabric (step 2) and compared with the AutoCad layer, only one sample (8.33 percent) was ranked 1 (i.e., containing > 90 percent of the parcel lines within +/- two feet of the lines in the control layer (Table 4)) and ten samples (83 percent) ranked 4 (i.e., < 50 percent of the parcel lines being within +/- two feet of the control layer lines (Table 4)). After applying the least-squares adjustment (see first adjustment, Table 4), the number of parcel lines that were within +/- two feet was somewhat improved. The number of parcels ranked 4 was reduced from 83 percent to 16 percent, 50 percent of the samples were ranked 3, and 25 percent were ranked 2 (Table 4). After evaluating each sample for adjustment performance and addressing any deficiencies or inaccuracies (see Table 5), a second adjustment was applied, resulting in 75 percent (9 out of 12 samples) achieving greater than 75 percent of the parcel lines falling within +/- two feet of the control layer lines.

Of the 12 samples that were adjusted, one sample (sample 3)

Table 4. Summary of the quality of the parcels prior to applying any adjustment, after the first adjustment was applied and again after revisions were made for each sample and a second adjustment was applied

Sample	Rank % Match Preadjust.	Rank % Match after 1st Adjust.	Rank % Match after 2nd Adjust.	Problem of Accuracy	Fix
1	4	3	1	Inadequate control (c)	Points added: 4
2	4	3	2	Disproportionate control (d), inadequate control (c)	Points added: 3
3	4	4	4	Bad parcel shapes (a)	Needs to be redigitized from plat
4	4	2	1	Disproportionate control (d), inadequate control (c)	Points added: 6
5	4	3	1	Disproportionate control (d), inadequate control (c)	Points added: 3
6	4	4	3	Disproportionate control (d), inadequate control (c)	Points added: 5
7	3	2	2	Disproportionate control (d), inadequate control (c)	Points added: 1
8	4	3	1	Disproportionate control (d), inadequate control (c), bad control (b)	Points deactivated: 1 points added: 2
9	1	1	1	No problem	
10	4	3	3	Disproportionate control (d), inadequate control (c)	Points added: 6
11	4	2	2	Disproportionate control (d), inadequate control (c)	Points added: 6
12	4	3	2	Disproportionate control (d), inadequate control (c)	Points added: 5

Table 5. Adjustment statistics for sample 3 and sample 9

Description	Sample 3	Sample 9
# Control Points	10	14
# Parcels	26	21
# Points	54	37
# Bearings	176	140
# Distances	176	140
# Unknowns	114	67
Redundancy	238	213
Bearings > Tolerance	3	0
Distances > Tolerance	9	0
Close Points Found	0	0
Line Points Found	0	0
Max. Easting Shift	-17.206	-9.885
Max. Northing Shift	-25.805 (745)	-10.016 (1060)
Avg. Easting Shift	0.123	-1.253
Avg. Northing Shift	-1.527	-0.051
Avg. of Coordinate Residuals	1.24	0.77
Std. Deviation Coordinate Residuals	7.77	4.51
Adjustment Rank	4	1
Comments	Did not converge or stabilize. Failed after 4 iterations.	
Number of Control Points	10	14
Number of Control Points Inside	1	4
Number of Control Points Outside	10	10
Lines within 2 Feet of Control Layer	14	35
Total Number of Lines	60	36
Percent Match	23.33	97.22

Figure 10. Images of sample 3 (lowest ranking) and sample 9 (highest ranking) before and after adjustment was applied

the original plats, joined to the fabric, and readjusted. No matter how many times a least-squares adjustment is performed, if the shape being adjusted is not at least representative of the space available, it will never reach an ideal solution.

A significant amount of improvement was made in the sample areas adjusted by adding additional control points and ensuring that they were well distributed inside of and around the boundary of the area being adjusted.

CONCLUSIONS

The parcel fabric data model provides a comprehensive way to manage cadastral information that can maintain historical parcel information in conjunction with detailed, survey information including geodetic coordinates. Once the cadastre has been created, it also can be continuously improved over time and efficiently associated with parcel-based layers, as illustrated by the successful achievement of the objectives set forth in this study. These include (1) developing a feasible workflow for converting existing data; (2) maintaining and improving the integrity of cadastre data over time; and (3) being able to integrate these data with related layers. This data model has provided Rapid City with the ability to improve its digital cadastre with a limited amount of resources. Understanding that care should be used when adjusting data of unknown or poor quality, it has been suggested to Rapid City that as long as the adjustments being made are checked against information of known good quality, this is a reasonable way to move forward and improve the quality of the existing data.

Land records information has historically been stored in GIS databases by individual components: points, lines, and polygons. One distinct weakness of this data model has been its inability to associate line and point features to the polygons they represented. There was also no efficient process or method that allowed for new

showed no improvement; four samples (1, 4, 5, and 8) improved from rank 4 to rank 1 (> 90 percent of lines within +/- two feet); two samples improved marginally (rank 4 to 3 (50 percent to 74 percent of lines within two feet)); and the remaining samples improved to within 75 percent to 89 percent of lines within +/- two feet. The reasons for these improvements and lack of improvement (e.g., sample 3) are summarized in Table 4. Overall improvements were possible by adding between one and six control points.

The lowest ranking area was sample 3 and the highest ranking area sample 9 (Figure 10 and Table 5). Sample 3's performance is the result of a bad shape. This is evidenced by comparing the parcels to the control layer, the adjustment solution not converging to zero, and by the high maximum northing shift in the adjustment statistics. To improve this area, the parcels should be re-input from

improved data to be incorporated, making it difficult to update property-dependent layers. In addition, distributing error for plat misclosures (Bunten 2008) also was challenging. The parcel fabric data model has addressed these shortcomings, resulting in a living land records system that is robust and more efficient to maintain and update.

Not only is it important to have a digital parcel dataset for assessing and collecting taxes and tracking land ownership, it also is becoming increasingly necessary that the accuracy and accessibility of land records information be improved for better resource management (Folger 2009), national security (Enemark 2010), critical infrastructure (Harper 2006), and emergency response efforts (Binge 2010). As pointed out by Brown and Moyer (1989), land is one of the most fundamental resources and, historically, records of this resource have been poor. However, as growth and development continue to occur, restricting the availability and challenging the resilience of this resource, having up-to-date and accurate information will be critical for the decision-making process. Craig and Wahl (2003, p. 95) contend that by having accurate spatial representations of land in a GIS, "the decisions about the locations of improvements and resources on the land will not be subject to costly errors and assumptions." One example of a community striving to improve the management of its land resources by developing a seamless parcel dataset is highlighted by Bunten (2008). The city of Duluth, Minnesota, embarked on a five-year project to "actively try to better manage development, its infrastructure and protect the natural environment, including the Lake Superior watershed" (Bunten 2008). This project was undertaken prior to the introduction of the parcel fabric data model and some of the challenges of working with land records information as individual components (points, lines, and polygons), as highlighted above, were encountered. The workflow developed during this study could easily be applied by a municipal organization, such as the city of Duluth.

The findings in this study reveal that one of the biggest challenges in migrating to the parcel fabric is preparing and loading existing data. The workflow developed during this study provides a means for systematically finding and addressing some of these pitfalls, which will result in more efficient implementations. The accuracy assessment presented in this study also provides users with a means for identifying problems when applying adjustments in the parcel fabric and outlines steps that can be taken to correct these issues.

For several decades, there have been voices defending the
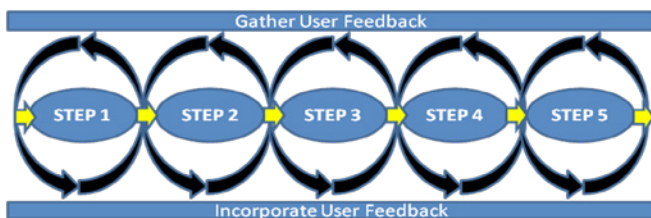
need for a nationwide cadastre in the United States (Foster 2008). While this has not been achieved to date, there have been successful statewide cadastres built, which is a step toward the goal of developing a national seamless parcel database. One such example is the state of Montana where the average annual benefit of having accurate accessible land records information is in the million-dollar range (Zimmer 2007). This example highlights the cost savings and efficiency realized by having an accurate, seamless dataset of land records. Countrywide digital seamless cadastral coverages of survey-grade accuracy also have been successfully developed. One such example is in New Zealand. Land Information New Zealand (LINZ) is an online seamless parcel data system that provides government officials, surveyors, and the public with more than 150 years of titles, survey marks, plans, etc., resulting in a significant increase in efficiencies for title research, land transfers, and filing of certified documents by surveyors. LINZ is supported by the New Zealand Institute of Surveyors and New Zealand Law Society (Richardson 2008). As more organizations adopt a common data model for storing land information, such as the parcel fabric, the effort of moving the United States toward a National Cadastral Dataset, as provided for in the National Spatial Data Infrastructure, will be strengthened.

Historically, surveyors have been remote from the GIS industry because GIS cadastral coverages were not representative of the precisions maintained by surveyors (Harper and Lee 2008). However, limitations in hardware and software that existed previously have largely been overcome. "Survey accuracy in a cadastral database encourages a mutually beneficial environment for both surveyors and GIS professionals" (Harper and Lee 2008). The development of a national parcel database would provide an opening for surveyors to be leaders in geospatial technology by viewing their work as a societal resource rather than a proprietary asset (Jones 2010).

FUTURE WORK AND RECOMMENDATIONS

The workflow that was developed during this study was an iterative process that included significant involvement from the end user (Figure 11) at each step, resulting in a process that can be implemented immediately. In fact, the workflow developed here currently is being used by the city of Rapid City to convert existing cadastral data to the parcel fabric. Because the workflow is generalized and quite scalable, it can be implemented elsewhere with other datasets for the principle requirements are the same (i.e., develop a framework (step 1); prepare the data (step 2); adjustment of the data (step 3); quality checking through accuracy assessment (step 4); and adjustment of associated layer(s) (step5)). The workflow can be adopted by both large and small organizations managing land-records information in both the public and private sectors. The applicability of this workflow is further supported by the response received at the GIS in the Rockies Conference 2011, where this work was presented. Representatives from a variety of sectors, including local governments, utility

Figure 11. Interactive and iterative process used for testing workflow usability

companies, the software vendor (ESRI), and private corporations, all expressed interest in the workflow that was developed.

Even though the workflow created during this study can be widely applied, the next logical progression of work to be conducted on this project is developing a subsequent workflow for Rapid City to identify specific processes for handling daily tasks once the legacy data has been migrated to the parcel fabric. Some of these include integration of new land transactions into the fabric, adjusting parcels to control points, incorporating newly acquired control points, refining cartographic elements (e.g., dimension annotation, parcel labels, etc.), and publishing the parcels dataset via a Web-mapping interface for end-user consumption.

Acknowledgments

About the Authors

Linda M. Foster, PLS, GISP, graduated from Pennsylvania State University with a Master's in geographic information systems and a Bachelor of Science degree in geological engineering. She is currently the GIS Manager for Ferber Engineering Company, Inc., in Rapid City, South Dakota.

Corresponding Address:
Ferber Engineering Company, Inc.
729 E. Watertown Street
Rapid City, SD 57701
Phone: (605) 343-3311
E-mail: lindafoster@ferberengineering.com

Dr. Justine Blanford is a researcher at the GeoVISTA Center in the Department of Geography and an instructor and adviser in the online Master's of GIS program at the Pennsylvania State University. She earned her PhD at Imperial College, London, UK. Her research interests include GIS and spatial analysis, spatial and temporal modeling, movement analysis and utilizing social media for crisis management.

Corresponding Address:
Geography Department
The Pennsylvania State University
302 Walker Building
University Park, PA 16802
E-mail: jib18@psu.edu

Notes

1. An individual parcel of land on which the identification of land rights resides (Enemark 2010) and an official register of the value and ownership of a parcel of land used in assigning taxes (Robillard et al. 2011).

2. The NMAS states that "for maps on publication scales larger than 1:20,000, not more than ten percent of the points tested shall be in error by more than 1/30 inch, measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch" (USGS 1947).

3. A standard publication scale for cadastral mapping is in the 1:1000–1:1200 range (FIG 2009; Kennedy and Ritchie 1982), translating to an accuracy of 90 percent of all measureable points/lines falling within +/- 3 feet to +/- 3.33 feet (Foote and Huebner 1995) and tested by comparing to corresponding positions as determined by surveys of higher accuracy (USGS 1947).

References

Bhowmick, A., N. Bodnar, D, Farmer, P. Tirunagari, and D. Van Pelt. 2008. Cadastre management the GIS way. Professional Surveyor Magazine 28:8. Retrieved April 8, 2001, from http://www.profsurv.com/magazine/article.aspx?i=2198.

Binge, M. L. 2010. Surveying GIS: The national cadastre. Point of beginning. Retrieved April 12, 2011, from http://www.pobonline.com/Articles/Column/BNP_GUID_9-5-2006_A_100000 00000000857492.

Brown, P., and D. Moyer. 1989. Multipurpose land information systems: The guidebook. Federal Geodetic Control Committee. Retrieved September 30, 2011, from http://www.ngs.noaa.gov/FGCS/tech_pub/Guidebook1of3.pdf.

Bunten, R. 2008. Duluth digital plat and parcel development for Lake Superior watershed program. Minnesota's Lake Superior Coastal Program. Retrieved September 26, 2011, from http://www.scribd.com/doc/28107114/Duluth-Digital-Plat-Parcel-Development-for-Lake-Superi or-Watershed-Protection-306-12-08.

Capobianco, K. M., and W. C. Mann. 2009. Landbase accuracy with cadastral editor. Government Engineering (Jan.-Feb. 2009): 20-21. Http://www.govengr.com/Articles Jan09/jacksonville.pdf.

Craig, B., and J. Wahl. 2003. Cadastral survey accuracy standards. Surveying and Land Information Science 63(2): 87-106.

Denver GIS Division. 2011. Telephone communication with Doug Genzer. City of Denver, Colorado.

Elfick, M., and T. Hodson. 2006. Managing cadastral data in a GIS. Shaping the change. XXIII FIG Congress Proceedings. Retrieved November 14, 2013, from http://www.fig.net/pub/fig 2006/papers/ts60/ts60_05_elfick_hodson_0739.pdf.

Enemark, S. 2010. Land governance: Responding to climate change, natural disasters, and the millennium development goals. Surveying and Land Information Science 70 (4): 197-209.

ESRI and J. Kaufmann. 2004. ArcGIS cadastre 2014 data model vision. Http://www.esriro.ro/industries/cadastre/pdf/nc_2014.pdf.

ESRI. 2010. Curves and lines. SDK sample add-in. Retrieved June 11, 2011, from http://resources.arcgis.com/gallery/file/arcobjects-net-api/details?entryID=3F5A0E12-1422-2418-A0DB-BA23CB137515.

ESRI. 2011. Desktop 10 help. ArcGIS resource center desktop 10. Retrieved May 18, 2011, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html.

Folger, Peter. 2009. Issues regarding a national land parcel database. Congressional Research Service. Retrieved October 19, 2011, from http://www.fas.org/sgp/crs/misc/ R40717.pdf.

Foote, K., and D. Huebner. 1995. Error, accuracy, and precision. The Geographer's Craft. Retrieved from http://www.colorado.edu/geography/gcraft/notes/error/error_f.html.

Foster, R. 2008. A national cadastre for the U.S.? Point of beginning. Retrieved June 12, 2011, from http://www.pobonline.com/Articles/Column/BNP_GUID_9-5-2006_A_100000000 000004 76418.

Genzer, D., and P. Tessar. 2011. *Migrating the DenverGIS Cadastral Coverages to the Parcel Fabric. Presented at the 2011 ESRI International User Conference, San Diego, CA, July 12, 2011.*

Geodata. 2006. Cadastral management solutions: A strategy for a gradual upgrade to a survey accurate GIS cadastral database. Geodata Information Systems. Retrieved June 5, 2011, from http://www.geodata.com.au/downloads/Gradual%20Upgrade%20Strategy %20Dec%2006.pdf.

Goodchild, M. 1992. Sharing imperfect data. NCGIS Initiative 9: Institutions Sharing Spatial Information Meeting. Retrieved June 22, 2011, from http://www.geog.ucsb.edu/~good/papers/228.pdf.

Harper, I. 2006. Survey accurate parcel networks as the base layer for enhanced GIS. ESRI Survey and GIS Summit—Paper 2436. Retrieved June 4, 2011, from http://www. geodata.com.au/downloads/ESRI%20UC%20Paper%202436%20-%20Ian%20 Harper.pdf.

Harper, I. and R. Lee. 2008. New GIS-based cadastral precision efficiencies. GIS development. Retrieved November 15, 2013, from http://www.geospatialworld.net/ magazine/MArticleView.aspx?aid=19293&Itemid=416.

Huong, D. 2010. The application of ArcGIS cadastral frabric model for cadastral database management. Presented at the International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences. Http://wgrass.media.osaka-cu.ac.jp/gisideas10/viewpaper.php?id=383.

International Association of Assessing Officers (IAAO). 2009. Standard on digital cadastral maps and parcel identifiers—2009. Retrieved August 14, 2011. from http://www.iaao.org/uploads/StandardonManualCadastralMaps.pdf.

International Federation of Surveyors (FIG). 2009. Cadastral template. A worldwide comparison of cadastral systems. Retrieved August 12, 2011, from http://www. fig.net/cadastraltemplate/fielddata/d1.htm.

Jones, B. 2010. National land parcel database and surveyors. Professional Surveyor Magazine 30(10). Retrieved April 9, 2011, from http://www.profsurv.com/magazine/article.aspx?i=70809.

Kaufmann, J., and D. Steudler. 1998. Cadastre 2014—A vision for a future cadastral system. Paper and Presentation. at Technical Session 7, XXI FIG Congress, Brighton, July, 1998, http://www.fig.net/cadastre2014/.

Kennedy, E. A., and R. G. Ritchie. 1982. Mapping the urban infrastructure. Paper presented at the Fifth International Symposium on Computer-Assisted Cartography, August, 1982. Retrieved August 12, 2011, from http://mapcontext.com/autocarto/proceedings/auto-carto-5/pdf/mapping-the-urban-infrastructure.pdf.

Konecny, G. 2011. A geocoded cadastral fabric as precondition for a sustainable land management system. GIS Ostrava 2011 Symposium Proceedings. Retrieved May 4, 2011, from http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2011/sbornik/papers/Konecny.pdf.

Moran, T., M. Herrmann, and T. Hodson. 2008. *The ArcGIS cadastral fabric data model. Presented at 2008 ESRI International User Conference, San Diego, CA, August, 2008.*

Rapid City GIS Division. 2009. Rapid City Corporate Limits [metadata]. City of Rapid City, South Dakota: GIS Division.

Richardson, K. 2008. Out with paper. Professional Surveyor Magazine 28(9). Retrieved April 9, 2011, from http://www.profsurv.com/magazine/issue.aspx?i=144.

Robillard, W., D. Wilson, and C. Brown. 2011. Evidence and procedures for boundary location. Hoboken, NJ: John Wiley Sons, Inc.

U.S. Geological Survey (USGS). 1947. National map accuracy standards. Retrieved August 10, 2011, from http://nationalmap.gov/standards/pdf/NMAS647.PDF.

Zimmer, R. J. 2007. Surveyors working to improve GIS. The American Surveyor. Retrieved April 9, 2011, from http://www.amerisurv.com/content/view/3764/150.

UPCOMING URISA® CONFERENCES

February 24-27, 2014
GIS/CAMA Technologies Conference
Jacksonville, Florida

April 14-16, 2014
CalGIS 2013 Conference
Monterey, California

May 5-9, 2014
URISA Leadership Academy
Calgary, Alberta Canada

September 8-11, 2014
GIS-Pro 2013: URISA's 52nd Annual Conference
New Orleans, Louisiana

October 13-17, 2014
URISA Leadership Academy
Louisville, Kentucky

www.urisa.org

Your Decisions Affect Theirs

Government decisions affect more than 300 million Americans a year. With Esri® Technology, you can connect with your entire constituency. Esri helps you demonstrate accountability, foster collaboration, and make the effective decisions that keep your constituents happy.

Learn more at esri.com/urisanews

esri®