Geography Compass 3/2 (2009): 698–726, 10.1111/j.1749-8198.2008.00209.x

Parcel Data for Research and Policy

Steven M. Manson*, Heather A. Sander, Debarchana Ghosh, J. Michael Oakes, Myron W. Orfield, Jr, William J. Craig, Thomas F. Luce, Jr, Eric Myott and Shipeng Sun
University of Minnesota

Abstract

Parcel data, or information about individual plots of land, may be used to examine a broad range of social and environmental issues. While analog parcel information has long been available, the move towards digital georeferenced data offers a more readily available means of using detailed structural and land use information. Parcels map onto useful units of analysis such as individuals and households, and serve as the foci of policy institutions tasked with functions such as taxation, schooling, zoning, or public health. The promise of digital parcel data is offset by challenges related to format, availability, maintenance, quality, augmentation, and confidentiality. We examine the use of digital parcel data in research and policy with special reference to land use, public health, education, and environmental applications. We also make recommendations for improving and using parcel datasets.

Introduction

Parcel data, or information about individual plots of land, are increasingly important to addressing a variety of research and policy concerns. These data relate to characteristics that range from the simple – such as location and size of the parcel – to the complex – such as features of structures on the land or information about the people who live there. The parcel is a convenient unit of analysis for many issues because parcels are the primary vehicle by which land is developed, used, exchanged, and taxed. Parcel data also give insight into the behavior of decision-makers ranging from individuals such as home owners, realtors, and developers to families and households who define many aspects of society. Parcels are also a useful unit of analysis for understanding the actions of organizations concerned with an array of issues, such as taxation, housing conditions, agriculture, marketing, infrastructure and service provision, zoning, health, education, poverty, and economic development. Overall, parcel data support high-resolution analysis that lies at the heart of new research directions (Longley 2003) and exciting policy opportunities (Treuhaft and Kingsley 2008).

There are a growing number of calls for fine-scaled analyses with parcel data to examine issues in human, natural, and human-environment systems.

Calls for fine-scaled analyses with parcel data and associated information have been issued in fields including demography (Entwisle 2007), land change (Irwin and Geoghegan 2001), hydrology (Arthur-Hartranft et al. 2003; Rodriguez et al. 2005), forestry (Brown 2003; Mansfield et al. 2005), transportation (Cervero and Duncan 2007), real estate (Cunningham 2006), public health (Dearwent et al. 2001; Miranda et al. 2007), remote sensing (AlGarni 1996; Liverman et al. 1998), economics (Geoghegan et al. 1997), and ecology (Olivera and DeFee 2007; Troy and Wilson 2006). This increased focus on fine-scaled analysis reflects the growing interest in the use of spatial data, methods, and theory more generally (Gewin 2004; Longley 2000).

Digital parcel data have great potential to improve analyses conducted in a variety of areas, but many hurdles remain in their use. Both parcel attribute data (information about the parcels) and parcel geometry (the parcel location or borders) have long existed in analog format, but digital parcel data consist of parcel geometry and attributes related to their sale, assessed value, and structural characteristics. Unfortunately, this growing availability is offset by numerous issues related to cost, availability, quality, completeness, maintenance, and confidentiality that make some parcel datasets difficult to use.

We examine the use of parcel datasets in research and policy in a variety of contexts, but for the purposes of exposition, we examine cases within the specific areas of land use, environmental applications, and public health and education policy. These cases center on issues such as lot-by-lot changes in land use and its attendant impacts (Carruthers 2002; Fox et al. 2003; Theobald 2001); house-by-house differences in access to health and educational opportunities (Berke et al. 2007a; Craig 1998; Lee et al. 2006; Miranda et al. 2007; Moudon et al. 2007; Roemmich et al. 2007); and provision of ecosystem services like natural habitat and water purification that vary at fine scales (Atasoy et al. 2006; Conway 2005; Rogers and DeFee 2005). We draw on international research and most of our findings are applicable globally, but we focus on the United States for much of our analysis, and in particular the Minneapolis-St. Paul metropolitan area of Minnesota, hereafter referred to as the Twin Cities Metropolitan Area (TCMA). We then identify key areas that benefit from parcel data and discuss the shortcomings of current parcel datasets with respect to research and policy. We also make recommendations for improving and using parcel datasets in the future. Finally, the article concludes with a resource section that guides the interested reader to range of articles, books, and websites that detail many aspects of accessing and using parcel data.

Parcel Data

Parcel data records treat discrete plots of land as their fundamental enumerative units. Digital parcel datasets are increasingly being maintained by local, regional, and state governments for a range of uses that typically include recording property taxation information, land use, and sales information. These data are also maintained by private groups, such as realtors, insurance companies, and real estate multiple listing services (MLS). The content of parcel datasets varies with region and maintainer. Some parcel datasets consist only of geometry and identification numbers, while others associate a large number of attributes with each parcel.

ACQUISITION

There are several routes to acquiring parcel data, with the caveat that parcel information is most useful when it is digital and georeferenced (e.g. having spatial coordinates for the parcel's location on the surface of the earth) either as a single parcel centroid or as a digital version of the parcel boundaries (Von Meyer 2001). There are alternatives to digital georeferenced parcel data, however. At the minimum, most local assessors maintain property and taxation records that may be used as surrogates for parcel map data. In this case, the parcel data most often exist as records tied to street addresses or lot identifiers. These in turn may be linked to paper plat maps, which indicate the boundaries of individual properties, and then manually surveyed or georeferenced by comparing the plat maps to known geographical features, such as a road map or survey networks, or using address-matching techniques.

To construct a history of land use on individual parcels, one can survey their owners or physically examine tax assessments and plat maps, but these approaches are limited because even midsized cities and counties can have tens of thousands parcels or more. Maps, remotely sensed imagery, or aerial photography may also be used to assess the developmental state or land use of many more locations, but these sources do not readily distinguish among individual parcels or directly indicate past land use beyond what is portrayed, and present land use or structural characteristics can only be inferred from what is seen from above (for case studies and best practices, see Huxhold et al. 2004; Stage and von Meyer 2006a; Treuhaft and Kingsley 2008).

Georeferenced digital parcel data are becoming increasingly available from government and private sources. Such data were relatively rare until recently, but the growing availability and usability of geographic information system (GIS) technology has allowed governments to more easily create and maintain digital parcel data, chiefly for tax assessment and collection (Treuhaft and Kingsley 2008). Availability of digital georeferenced parcel datasets varies by jurisdiction Stage and von Meyer (2006b) estimate that there are roughly 152 million land parcels in the United States, of which an estimated 70% exist in a spatially referenced digital format, mostly in urban areas (see also NRC 2007; Treuhaft and Kingsley 2008). Minimal coverage exists for most rural areas. In areas where such data do not exist, researchers may be forced to use paper tax records or plat books to construct their own spatially referenced digital parcel dataset as described above. In two related studies that focused on land use and ecosystem services, researchers created a spatial parcel dataset from survey and taxation records, and then used these data to map land use change to establish the relationship between land development and the impairment of watershed ability to prevent flooding (Olivera and DeFee 2007; Rogers and DeFee 2005).

Internationally, parcel data availability is limited and quality varies widely. Rajabifard et al. report on the state of parcel and cadastral systems around the globe, noting that only ten countries have compete nationwide cadastral systems, namely, Belgium, Brunei, Czech Republic, Denmark, Germany, Hungary, South Korea, the Netherlands, Sweden, and Switzerland (2007, 285). Others are like the United States in having some national standards but not national coverage, or national coverage but not necessary cadastral data as such; a good case in point is the United Kingdom, which has created a business-oriented national mapping program, the Ordnance Survey's 'OS MasterMap' with addresses and many parcels for the entire country (Brown 2004). More broadly, there are no internationally accepted standards for developing or assessing land administration systems (Steudler et al. 2004), despite these systems arguably being essential for sustainable economic development and effective governance (Williamson et al. 1999).

In order to illustrate the state of parcel data, we canvassed a variety of jurisdictions to establish the range of available data. We used as a starting point the recent NRC report, National Land Parcel Data: A Vision for the Future (2007) and then broadened our search via Internet and literature review (Table 1). A useful secondary result of this search was to establish that there is often a gap between the purported or reported data quality and what was actually available. For example, the NRC report details the state of the art of parcel data for many areas, but we found that in the time since publication, a number of jurisdictions identified as offering data sets changed the data or delivered different amounts than indicated. We examine several of these cases below, but refer the reader to the NRC report for an in-depth look at parcel data in the United States and the Resources section and Rajabifard et al. (2007) for more information on international data.

To examine issues of data quality more deeply, we delved into the parcel dataset for the TCMA, composed of Minneapolis; its sister city St. Paul, the state capital; and surrounding suburbs and rural communities. The TCMA was one of the first jurisdictions nationwide to collate parcel information in GIS format for a large, multi-county metropolitan region. This 7700-km^2 seven-county area is the economic hub of a multi-state region. It is home to 2.8 million people and is forecasted to top 3.5 million by 2020. The region's 272 local units of government – including 188 townships and cities – operate within a comprehensive regional

Table 1. Representative United States digital parcel records

Name: AK County Assessor Program Parcel Points (http://www.geostor.arkansas.gov/Portal/index.jsp)
Format: Points
Ownership: PID, owner, address
Tax: Assessed value of improvements, land, and total; assessment date
Sale: None
Parcel: Acres
Structural: None
Land use: None
Neighborhood: Neighborhood, school district
Spatial: Covers approximately 1/3 of state
Temporal: 01/01/2004–10/10/2005

Name: Dane County, WI (http://www.co.dane.wi.us/lio/)
Format: ArcGIS polygon shapefile
Ownership: Parcel identification number, deed restrictions, address
Tax: Easement type, tax district, assessed land and improvements values, previous year's land and improvement values, tax district
Sale: None
Parcel: Area, perimeter
Structural: Unit type, conditional use permits
Land use: Zoning category, water
Neighborhood: Zoning category, school district, plat/subdivision
Spatial: Dane County, Wisconsin
Temporal: From April, 2004, current to 2 months

Name: NY State Real Property Data (http://www.nysgis.state.ny.us/gisdata/index.cfm)
Format: Points by county; polygons for some counties
Ownership: Owner, address, deed location (book and page), owner type, PID
Tax: Easements, land assessment, total assessment, homestead code, roll section
Sale: None
Parcel: Acres, depth, front feet, property class
Structural: property class
Land use: Property class
Neighborhood: School district
Spatial: All but nine counties in New York State
Temporal: Annual datasets for 1996, 1998, 1999, 2000, 2004, 2005, 2006

Name: Portland OR Regional Tax Parcels (http://www.portlandonline.com/omf/index.cfm?c = 25779)
Format: ArcGIS Polygon Shapefile
Ownership: Parcel identification numbers, owner's name and address
Tax: County tax/levy code, Oregon Department of Revenue land use fields, tax valuation years, tax and land and total market values for two years, account status, responsible jurisdiction
Sale: Sale data, sale price
Parcel: Area, width, front footage
Structural: Year built, square footage of residence, # bedrooms, # floors, # units
Land use: Land use categorization
Neighborhood: none
Spatial: Portland and surrounding three counties
Temporal: From 1990 onward; updated weekly

Geography Compass 3/2 (2009): 698–726, 10.1111/j.1749-8198.2008.00209.x

Table 1. *Continued*

Name: Delaware County, OH (http://www.dalisproject.org/)
Format: Polygons
Ownership: PID, name, address
Tax: Tax district; class; market and taxable values of land, property, and total; annual tax
Sale: Sale dates and amounts for two transactions
Parcel: Acres
Structural: Condo. (Y/N); year built and remodeled; numbers of bedrooms, family rooms, full/half bathrooms, rec rooms, and total rooms; grade; height; air conditioning; number of fireplaces, heating type; finished area; number of units; basement (Y/N); crawlspace (Y/N); attic (Y/N); garage type
Land use: Use class
Neighborhood: Subdivision name, neighborhood code, school district
Spatial: Delaware County, Ohio
Temporal: Unknown

Name: Maryland (http://gis.mt.gov/)
Format: Polygons
Ownership: PID, name/address, deed reference
Tax: Exempt status and class; dates of last inspection and assessment; full market land, improvement, and total values
Sale: Seller name; deed info.; sale type & date; partial or total transfer indicator; down payment; mortgage; sale land, improvement, & total values; vacant/improved indicator for sale time; improvement value fields; capitalized ground rent

Name: Twin Cities MN Regional Parcel dataset, (http://www.datafinder.org)
Format: Polygons and points by county
Ownership: Name, address
Tax: Estimated market values of land, buildings, and total; tax capacity; total tax; special assessments; tax exempt status, exempt uses (4)
Sale: Sale date, sale value
Parcel: Deeded and polygon acres
Structural: Dwelling type, home style, finished square footage, garage, garage square footage, basement, heating, cooling, year built, number of units
Land use: Use type; multiple uses; green acres status; open space indicator; agricultural preserve indicator, enrollment, and expiration dates
Neighborhood: School district, watershed district, plat
Spatial: Minneapolis-St. Paul seven county area
Temporal: 2002–2006 (annual), updated quarterly in current year

Name: Travis Central Appraisal District (TCAD) Parcels (http://www.ci.austin.tx.us/landuse/gis.htm)
Format: Polyogs
Ownership: PID
Tax: Tax exempt codes; land and improvement market values in 1990, 1995, and 2000; total market value in 1990, 1995, 1997, 1998, 2000, and 2002; tax rates
Sale: Year of purchase
Parcel: Area, perimeter

Table 1. *Continued*

Parcel: Property factor influence (e.g. commercial/industrial influence, type of income-producing property, historical, topography type, lot shape); lot area, width, and depth
Structural: Utility type(s); year built; construction grade and material; number of stories; dwelling type; recreational use indicator; square footage of structure; number of dwellings, dwelling units, and rooms on parcel; owner occupied indicator
Land use: Land use class; commercial / industrial / residential land use types
Neighborhood: Zoning, multiple zoning indicator, Census 2000 census tract and block group, critical area code (indicates conservation, limited development, and intensely developed areas)
Spatial: Statewide
Temporal: Updated annually

Name: Massachusetts Digital Assessors' Parcels (http://www.mass.gov/mgis/parcels.htm)
Format: Polygons (some municipalities may contain more/fewer attributes)
Ownership: Site owner (not distributed), parcel identification information, ownership interest (i.e. owner's land use right)
Tax: Total assessed value for land and structures, year of valuation
Sale: Last sale date and price, lowest and highest house numbers on parcel
Parcel: Acreage
Structural: # living units, residential living area, building area
Land use: Land use
Neighborhood: None
Spatial: Municipalities: 46 (standards compliant) and 123 (not compliant)
Temporal: Current to 2003

Structural: Improvement code (indicates structure type), improvement (structure) square feet, construction year, units
Land use: Land use in 1990, 1995, and 2000
Neighborhood: None
Spatial: City of Austin, Texas
Temporal: Year 2000 (Parcel level land use datasets also available for 2003)

Name: Wake County, NC (http://www.wakegov.com/gis/default.htm)
Format: ArcGIS polygon dataset
Ownership: Owner name, owner address, property identification number, deed date and registration number
Tax: Building assessed value, land assessed value, tax district, billing class
Sale: Total sale price, sale date
Parcel: Deeded acres
Structural: Year built, number of dwelling units, utility type(s), building type/use, building style
Land use: Land class, building style
Neighborhood: Fire district, zoning
Spatial: Wake County, NC
Temporal: 1999 onward; updated monthly

planning framework, the Metropolitan Council, which is unique in the nation in terms of its authority. Established in 1967 in response to infrastructure issues caused by suburbanization, the state legislature broadened its mission in 1974, 1976, and 1994 to guide the efficient and sustainable growth of the region through land use planning, infrastructure development, and property tax revenue distribution. One advantage of this policy regime is a focus on collecting data at the regional scale and the ability to sponsor a concerted effort to gather and aggregate parcel data (Hayward and Mondale 2000). Individual counties collect parcel data for tax assessments, sometimes working with municipal assessors. These data are passed on to MetroGIS, which is part of the Metropolitan Council, which then makes it available to others (Johnson 2005).

While governments are the most common provider of parcel data, private companies increasingly maintain parcel datasets for their own use and for sale. Some real estate MLS, for example, maintain and sell parcel data, as do other private groups concerned with cadastral data, such as banks or land developers. Maintainers of these datasets often construct them by using assessor data from public parcel datasets as a base and then adding real estate sales listings and other data gleaned from private sources, such as infrastructure or utilities information. Parcel data are also frequently cleaned and augmented by private companies to produce databases of house addresses for use by businesses such as delivery companies or public sector organizations such as emergency services. These private parcel datasets therefore typically contain more detailed and higher quality data than government parcel datasets. Northstar MLS, a private company supporting realtors in the Twin Cities and across the state, maintains data for nineteen Minnesota and three Wisconsin counties (E. Newman, personal communication). This company purchases county assessor data and then improves it, correcting the errors it contains, and reconciling the different coding schemes for storing land use information used by different townships and counties. Northstar MLS then combines these parcel data with its own realtor information, which details a host of structural information (e.g. number of bedrooms, existence of air conditioning) and neighborhood characteristics (e.g. watershed, electoral district). This process thus results in a considerably more detailed dataset than that produced by MetroGIS.

The costs of producing and maintaining parcel data can be high. Government agencies face the need to recover costs, but the extent of this support arguably ranges from charging reproduction costs for data that have already been collected through to charging a premium to support activities tangentially related to the parcel data (Craig 2005; Johnson 1995). Los Angeles County, for example, originally charged thousands of dollars for information on 2.3 million parcels over 10,570 km^2. These charges were reduced to the cost of data reproduction, or about ten dollars, upon the advice of California's Attorney General and Department of Public Works to ensure the county met the requirements of the state's public records laws (Lockyer 2005). In the Twin Cities, in contrast, Northstar MLS pays about $43,000 per year for raw assessor data (two counties, Hennepin and Washington, account for 50% of this cost) plus $44,000 for ancillary GIS information (about one-third from counties, the remainder private) followed by several times that in data collation and enhancement costs. These costs raise the larger and thorny issues of how much public agencies should charge for data they collect with public tax dollars, the effects of restricting information on education and democracy (Haque 2001; Klinkenberg 2003), the need to develop institutions and provide incentives to individuals who create and share these data (Craig 2005), how agencies should evaluate the success of information dissemination (Georgiadou et al. 2006), and the need for best practices in handling parcel data (Stage and von Meyer 2006a).

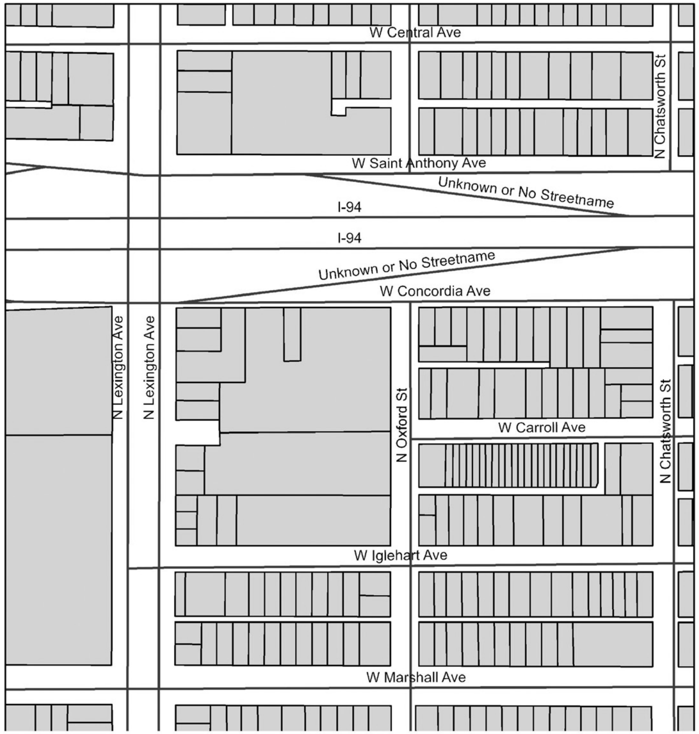
Despite their general high quality, the key drawback to private data is cost. The MLS dataset described above and other datasets like it contain more detailed parcel attributes than many publicly available datasets. This suggests that parcel datasets available from private companies would be superior to public parcel datasets as data sources for many forms of analysis and policy making. However, a large number of analyses that use parcel data are conducted by academic interests, community groups, or government agencies to whom local governments provide the data at reduced cost or free of charge, as is the case for the TCMA. The same is not true for privately produced parcel data which can be quite costly. In the case of Northstar MLS, for example, provision of the roughly million parcels for the TCMA is on the order of $50,000 under academic pricing, and far more for commercial interests. In the face of this expense, it is understandable that analyses are often executed using parcel data maintained by local and regional governments despite their sometimes lower information content and quality.

GEOMETRY

Digital parcel data in GIS format have geometry. The parcel dataset for the TCMA, for example, consists of parcel geometry as well as a detailed attribute table (Figure 1). In terms of sophistication, this geometry can range from points (indicating the parcel centroid or other location) to polygons outlining the parcel boundary to (more rarely) building footprints. Parcel location is important to homeowners and researchers alike. While detailed parcel data are useful in a range of analyses, parcel geometry alone is valuable to researchers as it provides spatial situational information that may not otherwise be available, such as texture, neighborhood composition, adjacency of houses or farmsteads, or local street networks. Related to this, even parcel point data can be used to derive accurate information on how the parcel lies in relation to other entities of interest, such as distance to the nearest highway or water body. The lack of detailed attributes

© 2009 The Authors
Journal Compilation © 2009 Blackwell Publishing Ltd
Geography Compass 3/2 (2009): 698–726, 10.1111/j.1749-8198.2008.00209.x

Georeferenced Parcel Data



Parcel Attriute Table

USE1_DESC	H	EMV_T	DWELL_TYPE	FIN_SQ
RESIDENTIAL	Y	124500	SINGLE FAMILY DWEL	972
RESIDENTIAL	Y	152100	SINGLE FAMILY DWEL	897
RESIDENTIAL	Y	158800	SINGLE FAMILY DWEL	897
COMM LAND & BL	N	24600		0
COMM LAND & BL	N	632200		0
RESIDENTIAL	Y	2900		0
RESIDENTIAL	Y	177900	SINGLE FAMILY DWEL	1214
RESIDENTIAL	Y	161700	SINGLE FAMILY DWEL	1112
RESIDENTIAL	Y	202100	SINGLE FAMILY DWEL	1907
RESIDENTIAL	Y	201200	SINGLE FAMILY DWEL	1790
RESIDENTIAL	Y	165000	SINGLE FAMILY DWEL	1092
RESIDENTIAL	Y	161900	SINGLE FAMILY DWEL	1380
RESIDENTIAL	Y	158800	SINGLE FAMILY DWEL	1332
RESIDENTIAL	Y	179700	SINGLE FAMILY DWEL	1482
RESIDENTIAL	Y	157600	SINGLE FAMILY DWEL	1040
RESIDENTIAL	Y	167000	SINGLE FAMILY DWEL	1195
RESIDENTIAL	Y	236200	SINGLE FAMILY DWEL	1800
RESIDENTIAL	Y	173400	SINGLE FAMILY DWEL	1104
RESIDENTIAL	Y	177300	SINGLE FAMILY DWEL	1116
RESIDENTIAL	Y	181700	SINGLE FAMILY DWEL	1667
RES DUPLEX/TRIP	Y	272100	DOUBLE DWELLING	1798
RES DUPLEX/TRIP	Y	230800	TWO FAMILY DWELLI	2110
RESIDENTIAL	Y	165400	SINGLE FAMILY DWEL	1040
RES 1 UNIT	N	154200	SINGLE FAMILY DWEL	1040
RES 1 UNIT	N	190200	SINGLE FAMILY DWEL	1788
RESIDENTIAL	Y	166100	SINGLE FAMILY DWEL	1330

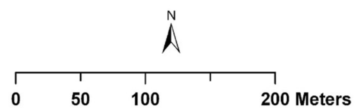


Fig. 1. Portion of the digital parcel dataset for St. Paul, Minnesota, illustrating georeferenced parcel data and accompanying attributes.

within a parcel dataset therefore does not necessarily indicate poor quality because such datasets may be in the early stages of development or are explicitly designed so that they might easily be linked to other GIS thematic layers.

ATTRIBUTES

Parcels provide a fine-scaled unit of analysis for both *in situ* attributes and characteristics that relate to their larger context, or in other words, site and situation. The former describe characteristics that are inherent or specific to each parcel, such as tax, price, structural, and land use information. Parcel data can also be assigned characteristics that describe how they are situated within a larger social or environmental context, such as distance to parks or the school district to which they belong. The latter information is comparatively rare in public parcel datasets and is normally obtained or derived from other datasets by using parcel geometry or identifiers as a link. We explore these data acquisition methods later in this article.

The distinction between *in situ* and contextual attributes is illustrated by one of the most common forms of parcel-based land use analysis, hedonic pricing, which estimates the relative contribution of different aspects of properties to their sales prices and uses this information to determine their marginal economic values. Under the assumptions that individual property buyers maximize their utility subject to budget constraints, that prices are in equilibrium, and that the area studied represents a single market, a hedonic model posits that the sale price of a home is a function of its characteristics, such as structure or lot size, in addition to a broader set of neighborhood, social, and environmental contextual characteristics ranging from school quality to the number of parks nearby (Freeman 2003).

Tax, Price, and Sales

Since parcel datasets are often maintained by local county or city assessor offices, they usually contain parcel identification numbers, property ownership, assessed property values, and property tax amounts. In some cases, parcel data may encompass a broader range of tax-related attributes, including specific assessed values for buildings, land, and total parcel and data related to tax exemptions, account status, and total taxes paid (Table 1). Tax-related attributes are often the primary reason for the existence of these datasets, so most of them report current assessed values, and tax-related attributes may be better developed and maintained than other attributes. The parcel dataset for Austin, TX, for example, contains attributes that provide a considerable amount of tax-related information, including assessed market values for multiple years, tax rates, and tax exemption information.

Some parcel datasets also contain sales information. When these data are provided, they normally refer only to the last sale transaction, sale date, and price. The parcel datasets for the TCMA and for Wake County, North Carolina, for example, report this basic information. Some jurisdictions provide more detailed sale information. Delaware County, OH, for example, provides sale date and value information for the last two sales of each county parcel and Maryland's parcel data contain a great deal of sales information including sellers' names, sale types, and mortgage and down payment amounts. Thus, sales information varies considerably by jurisdiction.

Not all parcel datasets include tax information, but parcels can be used as a link to this information. Some information of interest was missing in nearly all studies and datasets we reviewed, and, as a general principle, parcel geometry or street address must be used to extract this information from other datasets. In the simplest scenario, this may be accomplished by using parcel identification numbers or address matching to link to other datasets, such as sales transactions stored as tabular records (e.g. Bae et al. 2007; Geoghegan et al. 1997; Sengupta and Osgood 2003; Tajima 2003). However, adding these data to parcels datasets raises issues related to spatial and temporal scale mismatches and differences in formatting and projections that may complicate analyses. These issues are discussed in greater detail later in this article.

Structure and Land Use

Most parcel datasets contain some information about structures present on parcels. Detailed land use data may include features like vegetation, pavement, and buildings. These in turn can be used to evaluate the environmental impacts of development on the environment, such as water quality or microclimate (Stone 2004; Stone and Bullen 2006). Ellis et al. (2006), for example, linked household survey responses to parcel data to examine the effect of land use and tree cover patterns on neighborhood satisfaction. They were able to make subsequent policy recommendations on how to use tree planting to mediate the negative effects of proximity to commercial land use on homeowner satisfaction. Some parcel datasets, such as that of the City of Austin, TX, report only information related to the types, sizes, and years of construction of one building on each parcel. Others report considerably more structural information. For instance, the Maryland parcel dataset reports on the number of buildings on a parcel, building usage, and the numbers of stories and rooms in buildings. At the extreme, Delaware County, OH, reports considerably more structural information than most other jurisdictions in its dataset. This includes year of remodeling, the number of rooms of five types, number of fireplaces, heating type, and the presence of crawlspaces, basements, and garages. Thus, as with sales information, considerable variation exists in the degree of structural data provided by different parcel datasets.

In addition to explicit structural information, parcel datasets often report land uses from which limited structural information may be inferred, but these data vary greatly in the amount and type of land use attributes they contain. Most parcel datasets contain at least some indication of the land uses present on a parcel. This may be explicitly stated or may be inferred from tax or structural information (e.g. single family dwellings versus commercial buildings or warehouses versus agricultural uses). The parcel dataset for the TCMA, for example, varies with county in how such data are reported with some counties reporting highly specific land uses and structural information and others reporting only land use type.

In summary, parcel datasets provide a range of valuable data for use in assessing linkages among phenomena, such as land change, policy, and health. Without fine-scaled parcel data, many kinds of studies would be impossible. Use of these data, however, may be problematic, and we describe related challenges below, such as the need to augment parcel attributes with data from other datasets at different spatial or temporal resolutions, in different formats, and with different projections.

Challenges in Using Parcel Data

The increasing availability of parcel datasets has facilitated a growing array of research and policy applications, but using these data is challenging in several respects. Chief among these issues is availability of these datasets, which varies widely, as does their maintenance. Even when available and regularly updated, parcel datasets vary in their quality between and within jurisdictions. Other issues center on how parcel data must often be augmented with data that are collected or reported at very different spatial or temporal scales. Finally, it is a relatively straightforward process to identify individuals associated with parcels, leading to confidentiality and privacy issues.

AVAILABILITY

The growing accessibility of parcel datasets has enabled researchers to conduct studies in a timelier manner as they do not need to construct these datasets themselves. Data acquisition and the effort required to construct datasets from paper records are expensive, and can be prohibitive for large areas with many parcels. Readily available parcel data at minimal or no cost – particularly to academic, governmental, and not-for-profit organizations – has made possible many studies that otherwise would have been difficult or impossible to conduct.

Although digital, georeferenced parcel datasets have become increasingly common over the last 15 years, they still are unavailable in many regions. As noted above, coverage tends to be limited in rural areas, which is unfortunate given that these data are valuable in understanding and addressing rural issues related to agriculture, natural disasters, urbanization pressure, and ecosystem service provision (NRC 2007). Parcel data, for example, can inform decision-making related to habitat conservation in rural areas, because it forms the basis for decisions related to the purchase of land on a parcel-by-parcel basis by providing data for fine-grained assessments of the cost of these parcels and their value to conservation (Shilling and Girvetz 2007; Straeger and Rosenberger 2007) or for estimating and predicting the conservation easement value and likelihood of land use conversion for developable parcels (Newburn and Berck 2006). Additionally, even when parcel data are available for a given area, they may vary in their content. The entire TCMA, for example, has parcel data, but their depth and breadth vary widely by jurisdiction, as do the systems used to organize parcels and report attributes. In this case, parcel-level analyses may be restricted to areas for which data are available and roughly equivalent given the significant data reconstruction or preprocessing required for analysis. As we note below, in areas for which parcel data are incomplete, it is necessary to augment these data with other sources and kinds of information.

In the long run, national availability of parcel data will very likely require some coordination among governmental agencies and private sector firms. Federal interest in a national cadastral database was expressed by the report *Need for a Multipurpose Cadastre* (NRC 1980), which predicted that development of a national database would start with local entities operating within federal guidelines. The Federal Geographic Data Committee (FGDC), a federal organization that coordinates national spatial data activities, was founded in 1990 and was soon followed by the establishment of the National Spatial Data Infrastructure in 1994, a broad framework to coordinate spatial data nationally. The FGDC helped develop standards for National Spatial Data Infrastructure, one of which was for cadastre, codified as the *Cadastral Data Content Standard for the National Spatial Data Infrastructure* (FGDC 2003). The 1994 federal guidelines were not enough, however, and only modest progress has made toward the 1980 vision (Craig 2005).

Reasons vary for this lack of progress, but it is caused primarily by the fact that most parcel data are collected and managed at the local level and data sharing therefore meets immediate institutional and practical needs through informal mechanisms (Harvey and Tulloch 2006). This has led many to conclude that 'the creation of a nation-wide parcel-level dataset will require the participation of local government, finance agencies including Fannie Mae and Freddie Mac, realtors, and market researchers' (NRC 2003). This sentiment is reflected in recent moves by companies, such as MLS (noted above), Navteq (www.navteq.com), and Zillow (www.zillow.com) to generate their own parcel data, almost always based on public data, for profit. The international surveying community also sees the need for private–public partnerships that combine the public roles of supervision and coordination with private capacity for customer service and flexibility (Kaufmann and Steudler 1998).

MAINTENANCE

One problem with the use of parcel data relates to the manner in which they are maintained and updated. Parcel data can be used to understand parcel characteristics other than price, such as timing of land change or the contributing factors to shifts among land uses, but such analyses require parcel data to explicitly detail the timing of changes in land use, ownership, prices, or structural characteristics (Irwin and Bockstael 2004). In some cases, it is possible to infer these transitions through changes in reported market value, property sale prices, dates of subdivision or construction reported in parcels datasets (Newburn and Berck 2006). Landis et al. (2006), for example, developed a model to assess potential infill housing opportunities in California. The used parcel attributes including value and presence of structures and sizes to determine if a parcel was vacant or not. However, in many cases these data are absent or are overwritten when subsequent changes occur.

The overwriting of data in the maintenance of parcel datasets is problematic as it makes longitudinal studies nearly impossible. For example, when a parcel is developed, its land use may simply be changed in the dataset's attribute table or, when a parcel is sold, its new sale price may replace any values from previous sales. This lack of history in parcel attributes complicates efforts to track and analyze changes over time and may limit the usability of a dataset for temporal analyses. Parcel attributes from previous points in time could be preserved, for instance, by maintaining annual versions of parcel datasets as is done in the TCMA, or by maintaining attribute fields for different time periods as Delaware County, OH, does for sale dates and values. Tracking changes over time enables the users of parcel datasets to conduct a broader range of analyses, but this tracking is not common because assessor and tax offices are generally more interested in maintaining current records than keeping 'out of date' information on hand.

QUALITY

One inescapable fact of spatial analysis is that all spatial data suffer from quality issues that are inherent to how these data are collected and stored. The FGDC identifies many dimensions of data quality, of which four are particularly germane to parcel data: attribute accuracy, logical consistency, completeness, and positional accuracy (FGDC 1998). These dimensions provide a useful framework for exploring quality issues in parcel data.

Attribute accuracy concerns how well data describe the characteristics of phenomena. Given the complexity of the parcel data and the real world objects they represent, it is unreasonable to expect that every parcel in a dataset will accurately describe the real world parcel's attributes. Attributes for some parcel features, such as number of bedrooms in a structure, can be stored with some exactness, but this does not guarantee that the

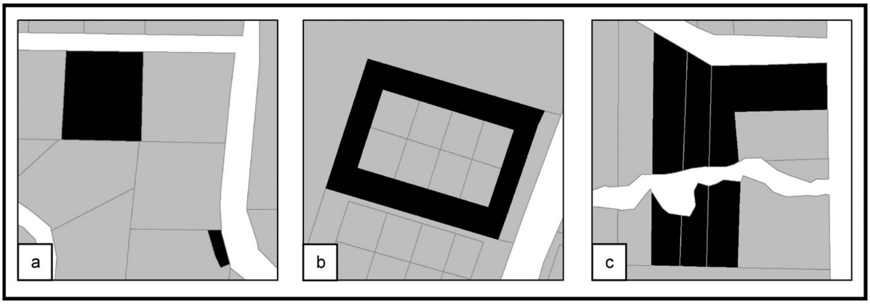


Fig. 2. Data errors in parcel dataset for Twin Cities of Minnesota: (A) depicts two parcels with the same PIN in different locations; (B) shows a parcel that is listed as residential single family and has attributes thereof, but is actually a road; and (C) shows three parcels that are listed as six parcels because they are divided by a stream.

number of bedrooms recorded for a parcel in the database match those in the real-world house situated on the parcel.

Logical consistency refers to the fidelity of data storage on the computer. Data consistency matters primarily from a technical standpoint. Parcel data are especially useful when they are topologically consistent, for example, which means they are stored in a way that allows the user to establish mathematical and spatial relationships among parcels, such as adjacency to one another, whether they border on the same road, or whether one is nested inside another. When parcel data are non-topological, the user must spend a good deal of time creating topological relationships in order to conduct many kinds of spatial analysis. More generally, given the large number of records found in parcel datasets and the various sources used to create them, it is not uncommon to find simple mistakes, such as parcels that are assigned incorrect information or non-parcels (e.g. roads and water bodies) being treated as parcels (Figure 2).

Positional accuracy describes the error between encoded spatial coordinates and their real-world counterparts. Some error is inherent to spatial data. Locations can be stored to the nearest millimeter, for example, but very few real-world measurements are collected with this precision. More prosaically, points are often used to approximate the locations of parcels, as they are in the parcel datasets for the states of New York and Arkansas. Given that parcel datasets themselves rarely contain all of the information of interest in an analysis, parcel geometry must be used to extract data from other layers (e.g. flood zones, wetlands, and rare species occurrences), as discussed earlier. Measurements based on points as opposed to parcel boundaries or based on imprecise parcel locations will contain errors in proximity metrics or in the values of variables extracted from other datasets, particularly for large or irregularly shaped parcels.

Beyond locational error, the greatest data quality issue that impairs the ability of researchers to use parcel data relates to its completeness, which can refer to both spatial data and attributes. The TCMA parcel dataset exemplifies these challenges. As noted above, these data are substantial and useful in many ways, and indicate the level of commitment and coordination necessary to produce a multi-county dataset. Nonetheless, this dataset is missing data for certain sub-areas, up to entire counties, for some time periods or kinds of housing. Similarly, analyses involving multiple counties require significant preprocessing to standardize data because counties interpret attributes differently, complete fields using differing terminology, or simply ignore fields. The general attribute 'Housing Style', for instance, is reported in several different ways. For example, Dakota County uses eight single family residential categories related to number of stories while Anoka County reports over one hundred residential and non-residential home style categories. Standardization of data entry across counties in this and other areas as well as the consistent use of all attribute fields would improve the usability of datasets for analyses. Again, through diligent effort and coordination among many agencies, the TCMA possesses one of most comprehensive digital parcel data datasets available in the United States, but this effort nonetheless faces data quality issues that reflect larger, persistent challenges in sharing geospatial data given their complexity and lack of common standards (Harvey et al. 1999).

Data quality issues arise largely from communities' inability to adequately fund or coordinate their efforts in collecting parcel data during the normal course of the taxation and assessment activities that provide most of the raw data that go into parcel datasets. These data in turn cannot be fully reported by counties for fear of entering incorrect data or, in some cases, because some data simply are not collected. Similarly, in many parcel datasets, attributes are omitted that are of potential interest, such as detailed information about the characteristics of structures on parcels that are maintained in other formats by assessor departments, but do not make it into digital versions of parcel data. The data quality challenges faced by users of the TCMA data are similar to those faced elsewhere, especially in terms of incomplete or inconsistently completed attributes, particularly in areas where multiple jurisdictions report parcel data (Table 1). These quality issues may weaken the utility of parcel data for many analyses.

AUGMENTATION

As noted above, few parcel datasets are equipped with all attributes necessary for most kinds of analysis. In this case, the data user must match parcels to other attributes at three different scales: below the scale of the parcel, at the same scale, and at larger scales. Matching parcels to other data at the same scale is usually straightforward, as when matching parcel identification numbers to sales information, but matching to lower or higher scales can be more difficult, as when linking parcels to individuals at lower scales or remotely sensed imagery or maps at higher scales. It is helpful to think of these broader contextual variables as comprising the three general categories explored below: distance metrics, focal characteristics, and zonal measures.

The prime form of matching parcels to lower-scale data involves linking a parcel to the people living there. This matching is usually performed by linking an owner identifier to a parcel identifier or a person's mailing address to the street address of a parcel as determined via other records, such as a personal interview or phone listing. Matters are made considerably more difficult when either the person or parcel does not have an exact location. In this case, the address of either a person or parcel can be estimated via street-based geocoding, which sites the location as a function of where an address falls within the range of addresses on its side of the road, but there is no guarantee that the address is correct. Although a growing number of data sets register the exact location of a given address by matching it against a database of known addresses, it is still common for a GIS to locate the most likely location based on how far the address is from the nearest street intersection or known address. This approximation can lead to errors in analyses based on participant locations, and in turn, in assessments that rely on these data. In an analysis of geocoding errors for assessing the exposure of children to traffic-related air pollution, for example, Zandbergen (2007) found that street-based address geocoding can create errors on the order of several hundred meters, making these data suspect for fine-scaled analysis (see also Burra et al. 2002). At this point, the researcher or policy maker is usually left to verify addresses by hand, drop them from consideration, or adopt different approaches such as trying to link a phone number to the address in order to interview the people there.

Conversely, using parcel centroids to locate participants greatly reduces mapping errors. Dearwent et al. (2001) examined the impact of these two methods (parcel centroid vs. street-based address geocoding) for locating subjects and found that noticeable differences existed between study participant locations identified using the two methods, which in turn resulted in misclassification of subject neighborhood characteristics. Parcel-level data may also be used to improve the sampling of subjects for health studies and may ensure that participants are distributed more evenly in a region and that specific regions of interest are adequately sampled (Lee et al. 2006). Parcel data therefore provide a promising avenue for locating subjects in space.

Parcel data are often linked to larger scale data, most often by using parcel geometry as a polygon overlay to extract or infer characteristics from larger scale data. Two common cases of such linking involve extracting data to parcels from larger enumerations (e.g. census boundaries, traffic analysis zones, or school districts) or using known characteristics of parcels for one area to make educated guesses about households in another area for which parcel data are not available. Analyses such as these that are conducted using data at both coarse and fine resolutions in combination will be limited by the precision of the coarser data, a problem termed scale mismatch, although there are a variety of techniques to consider under the general rubric of scale-dependent modeling. Miranda et al. (2002), for example, developed a model of childhood lead exposure by linking household health information (using blood tests to ascertain lead exposure) to parcel structural characteristics (e.g. year of construction and type of building) and aggregate data (e.g. census information and health care access). Other general approaches to scale mismatch or missing data include geostatistics and interpolation (Atkinson and Tate 2000; Tate and Atkinson 2001), spatial statistics (Anselin et al. 2004; Dungan et al. 2002; Waller and Gotway 2004), multi-level modeling (Bullen et al. 1997; Orford 2000; Vance and Iovanna 2006), and related solutions to ecological inference, where aggregate statistics are used to infer the characteristics of individuals, and the ecological fallacy, when these inferences are incorrect (Cho 1998; Dungan et al. 2002; King 1997; Robinson 1950; Wong 2004).

More prosaically, using parcel data to extract characteristics from other forms of data usually entails conversion among projections or formats, operations that necessarily generate errors. Most georeferenced data are projected, or are mathematically transformed to allow locations on a three dimensional sphere to be stored and analyzed as two dimensional data. Projection necessarily distorts data, changing properties such as angles and distance, as seen in the gross distortion of the poles in the commonly used Mercator projection. Converting between projections or reprojection can introduce further distortions, because the difference between two different projections must sometimes be approximated, especially at fine scales. Raster–vector conversion is similarly prone to error because the underlying data models are entirely different; small twists and turns of a vector line can be lost when moved to the blocky cells of a raster layer. Similarly, repeated conversion between vector and raster layers can introduce data degradation in every step (Peuquet 1981a,b). Many forms of spatial analysis can safely ignore these errors because the magnitude of error introduced by conversion operations is dwarfed by the inherent limitations of the data. For example, conversion errors on the order of meters are almost meaningless for data that is only accurate to hundreds of meters. Parcels boundaries, however, are often measured to the nearest meter or finer resolution, the same order magnitude as the errors caused by reprojection or raster–vector conversion.

Distance Metrics

Beyond site information, many studies also utilize parcel geometry in combination with other datasets to calculate distance from each parcel to a feature of interest – such as an amenity or a nuisance. Distance is often defined as Euclidean or 'as the crow flies' but other forms are also possible, such as Manhattan distance (distance along block faces), driving distance along a road network, or travel time by sidewalks. Hedonic pricing studies have calculated distances from individual parcels to land uses of interest, including open space (Acharya and Bennett 2001; Bolitzer and Netusil 2000; Irwin 2002; Lutzenhiser and Netusil 2001; Tajima 2003; Wu et al. 2004), wetlands (Doss and Taff 1996), forests (Mansfield et al. 2005), and agriculture (Ready and Abdalla 2005). Other work estimated the effects of distance to public transportation and transportation routes (Cervero and Duncan 2007) and central business districts (Lutzenhiser and Netusil 2001; Mansfield et al. 2005) on parcel price or land use. Positional errors in parcel or target feature location or mismatches between the scales of parcel and other datasets may lead to error in the calculation of these distances, as may differences in projection and data format. This error, in turn, may be propagated through models derived using them and may cause study conclusions to be inaccurate.

Focal Characteristics

Of growing importance to a wide array of analyses is treating the parcel as the 'focal' point of a small region for which some characteristic is calculated and then assigned to that parcel. The key here is that a common rule is used to define the region of interest for each parcel, such as a circular region with the parcel in the center or an area encompassing all locations within a given driving distance of the parcel. This region is then used as the basis for calculating some characteristic, such as the proportion of parkland or average air pollution that is then assigned to the focal parcel. While neighboring parcels will likely have similar focal characteristics given their physical proximity, each has its own associated focal area and so their focal characteristics will vary from one to another.

There are a variety of ways in which focal characteristics are used in research and policy. In the health context, there is much interest in individual participation in physical activities such as walking and biking or exposure to environmental risks such as pesticides, all of which vary as a function of housing density or transportation networks (Berke et al. 2007a,b; Moudon et al. 2005, 2007; Rull and Ritz 2003). Housing prices similarly vary as a function of access to parks, recreational facilities, or other services that in turn are a function of residential housing densities or transportation access in addition to the raw distance measures noted above (Cunningham 2006; Newburn and Berck 2006; Roemmich et al. 2006, 2007). Also of interest to a variety of researchers is the proportion of different land uses within a specified radius of a parcel (Acharya and Bennett 2001; Irwin 2002; Mansfield et al. 2005) and other measures of landscape condition such as the arrangement, diversity, and fragmentation of land uses around a given parcel (Geoghegan et al. 1997). Other studies considered remoteness from other development (Sengupta and Osgood 2003), air quality (Bae et al. 2007), greenness indices based on satellite imagery as a proxy to vegetation (Bae et al. 2007; Mansfield et al. 2005), and traffic levels (Cervero and Duncan 2007). Again, errors may occur in calculating these focal statistics as a result of differences in scale, format, and projection as well as errors in geometry that may impact the accuracy of analyses.

Zonal Measures

While distance metrics or focal characteristics vary incrementally from one parcel to another, large blocks of parcels can also belong to a single zone or region, with the result that there are large parcel agglomerations that have identical characteristics. Prime examples of zonal measures include assigning parcels to school districts (Bae et al. 2007) or zoning areas (Cunningham 2006; Irwin and Bockstael 2004) that relate to functions such as administrative jurisdiction, statistical reporting area, or environmental features such as a watershed or ecoregion. On a theoretical level, these zonations are interesting because they heavily influence parcel price or land use; two otherwise identical neighboring houses that happen to lie in two different school districts can have very different prices due almost entirely to the premium accorded to the better school district, and thereby access to education (Haurin and Brasington 1996). Perhaps the most common form of zonation involves applying the socioeconomic characteristics derived from census data to parcels, as these data are reported as arbitrary zonations ranging from small regions such as blocks and block groups to much larger census tracts (Acharya and Bennett 2001; Bae et al. 2007; Carrion-Flores and Irwin 2004; Geoghegan et al. 1997; Polimeni 2005; Waddell 2000). This approach has been used to approximate neighborhood income characteristics and racial composition in analyses of childhood blood lead levels (Miranda et al. 2002, 2007). As we describe above, there are a variety of challenges faced in assigning zonal characteristics to parcels given the likely mismatch in spatial scale among different sources of data.

CONFIDENTIALITY

There are a host of legal, moral, and political dimensions to geographical data, as with other kinds of data (Mugerauer 2000; Rhind 1996). We touched on some of these issues above, such as costs, challenges in reproduction and distribution, and quality issues. Particularly challenging problems with parcel data, however, hinge on confidentiality and privacy because location can act as a identifier to match across different databases in a way usually associated with names or identifiers such as social security, drivers license, or credit card numbers (Curry 1997). Researchers have long practice with protecting the confidentiality of data they collect, but the potential for location to let a third party infer characteristics of research subjects adds another layer of complexity that is the subject of a growing amount of research (Armstrong et al. 1999; VanWey et al. 2005).

Parcel data pose a special problem in that they are usually legally public and therefore do not fall under the aegis of confidentiality or privacy protection. Sharing data about individuals can cause anxiety, however, even if that data is public. This is especially true of parcel data because it will expose a person's home address to all users since parcel data nearly always contain the name of the owner and taxpayer. In response, some counties remove this information from the public version of their online parcel database to protect individual privacy. In the Twin Cities, for example, Dakota County originally removed all names at the request of some of their public officials, while following the law by giving out that information at public terminals in libraries and county buildings. Their rationale was that individuals with bad intentions would not likely expose their identities in such places. Anxieties dropped over time and the county subsequently restored names to their online parcel database. The county has not, however, made names one of the searchable fields online.

Conclusion

A broad range of analyses across many scientific, business, and policy realms increasingly rely on fine-scaled digital parcel data compiled by governmental and private groups. These data provide a variety of locational and attribute information and are available at scales that correspond to both the processes studied, such as the behavior of individuals and households, and the policies designed for them, such as taxation, economic development, schooling, or public health. Research and policy analyses conducted at this fine spatial scale may be readily combined with other kinds of data pertaining to individuals, such as personal health records or economic circumstance, or data on larger regions such as census information. Many forms of analyses would not be possible without parcel data, as seen in the many examples noted above.

However, along with the benefits provided by digital parcel datasets come several important drawbacks. Chief among these are data availability, maintenance, and quality. To these are added issues raised by having to augment parcels with other kinds of data and by confidentiality and privacy considerations. Any one of these challenges can negatively affect studies based upon parcel data. Addressing these issues requires diligence in all stages of data handling and analysis, from the moment of collection through to reporting final study results.

As parcel datasets continue to be made available in digital format and as their content and consistency is improved, a broader range of studies will become increasingly plausible, if not immediately possible. Members of the research community who utilize these parcel datasets, policy makers and community groups who rely on these data, and those charged with the creation and maintenance of parcel data should strive to improve the quality and coverage of parcel data and to attempt to ensure that the limitations of parcel data are clearly acknowledged and addressed in the future. Doing so will improve the usability of parcel data and the accuracy of studies and policies that use them. More broadly, the growing availability and use of parcel data is emblematic of the broader move towards spatial analysis in a broad range of fields (Fox et al. 2003; Gewin 2004).

Specific actions start with conducting research and implementing policies with digital parcel data that highlight their benefits and challenges. Researchers in part, but policy makers in particular, are well-positioned to treat parcel data as more than just another source of data and to actively contribute to the policy processes that identify needs for digital parcel datasets and mechanisms for their distribution. In addition to examining technical issues in sharing data, researchers must pay greater attention to the role of individual and institutional motivations in how data are produced, collated, and shared (Harvey and Tulloch 2006). Provision of parcel data ultimately relies on a range of individuals with the requisite idealism, enlightened self-interest, and engagement in a professional culture built on participation, cooperation, and trust (Craig 2005). Researchers and decision-makers have the responsibility and opportunity to nurture this professional culture while meeting their own research goals.

Resources

Several articles and books serve as excellent general resources on parcel data, particularly in terms of accessing these data and establishing best practices in using them. Cowen and Craig examine issues in developing cadastral data (2003). Huxhold offers a textbook that, despite its age, is still one of the best introductions to using parcel data and project management (1991); many of its themes are updated in discussing the use of GIS for local governance in Huxhold et al. (2004). Von Meyer offers another volume centered on GIS, but focusing on land records, including parcel data (2001). The NRC provided two detailed examinations of parcel data in the United States, particularly with respect to governmental roles and needs (1980, 2007). See Rajabifard et al. for information on international use and availability of parcel data (2007). Stage and von Meyer canvas a number of states to offer best practices in parcel management programs (2006a). Finally, Treuhaft and Kingsley offer an overview of parcel data for community development with six case studies (2008).

Interested readers are also directed to the following websites:

- Worldwide Comparison of Cadastral Systems: http://www.cadastraltemplate.org
- FGDC Cadastral subcommittee: http://www.nationalcad.org/
- Lincoln Institute of Land Policy: http://www.lincolninst.edu/
- MetroGIS: http://www.metrogis.org/

Acknowledgements

This work is supported by a Minnesota Population Center Faculty Proposal Development Grant, the National Aeronautics and Space Administration

New Investigator Program in Earth-Sun System Science (NNX06AE85G), and a Grant-In-Aid of Research, Artistry and Scholarship from the Office of the Dean of the Graduate School, University of Minnesota. The author gratefully acknowledges the assistance of Minnesota Population Center personnel and staff at the Regional MLS in St. Paul, Minnesota. Responsibility for the opinions expressed herein is solely that of the authors.

Short Biographies

Dr. Manson is an Associate Professor and McKnight Land Grant Professor in the Department of Geography at the University of Minnesota in the Twin Cities. Dr. Manson combines environmental research, social science approaches, and geographic information science to understand changing urban and rural landscapes. This work is part of his longer term research on global environmental change, decision-making, and understanding complex human-environment systems.

Heather Sander is a PhD student in the Conservation Biology Program at the University of Minnesota. Her research centers on the environmental and economic impacts of land use change.

Debarchana Ghosh is a PhD student in the Department of Geography, University of Minnesota. Her research interests include quantitative analysis, spatial statistics and modeling, GIS, health, and environmental geography.

Dr. Oakes is interested in quantitative methods, social epidemiology, and research ethics.

Myron Orfield is the Executive Director of the Institute on Race and Poverty and a non-resident senior fellow at the Brookings Institution in Washington, DC, in addition to being an Associate Professor of Law at the University of Minnesota Law School. He is an authority on civil rights, state and local government, state and local finance, land use, questions of regional governance, and the legislative process. He holds a joint appointment in Urban and Regional Planning in the Hubert H. Humphrey Institute of Public Affairs Graduate School. Professor Orfield holds a JD from the University of Chicago.

Dr. Craig is a member and past chair of the MetroGIS Coordinating Committee and vice chair of the Minnesota Governor's Council on Geographic Information. He was a member National Research Council committee that produced the 2007 report *National Land Parcel Data: A Vision for the Future*.

Thomas Luce focuses his work on state and local finance, metropolitan development, and intergovernmental relations. Co-author of a forthcoming book on regionalism in the Twin Cities, he is the Research Director of Ameregis, Inc. and the Institute on Race and Poverty at the University of Minnesota Law School. He has also been on the faculties of the Humphrey Institute of Public Affairs, University of Minnesota and the Department of Public Administration, Pennsylvania State University.

Eric Myott conducts and develops research at IRP. He has a bachelor's degree in Geography from the University of St. Thomas, St. Paul, and a Masters of Geographic Information Systems from the University of Minnesota.

Shipeng Sun is a PhD student in the Department of Geography at the University of Minnesota, Minneapolis. His research interests include intra-urban migration and agent-based modeling of land use change and urban growth.

Note

* Correspondence address: Steven M. Manson, Department of Geography, University of Minnesota, 414 Social Sciences, 267 19th Avenue South, Minneapolis, MN 55455, USA. E-mail: manson@umn.edu.

References

Acharya, G., and Bennett, L. L. (2001). Valuing open space and land-use patterns in urban watersheds. *Journal of Real Estate Finance and Economics* 22, p. 221.

AlGarni, A. M. (1996). Urban photogrammetric data base for multi-purpose cadastral-based information systems: the Riyadh city case. *ISPRS Journal of Photogrammetry and Remote Sensing* 51, pp. 28–38.

Anselin, L., Florax, R. J. G. M., and Rey, S. J. (eds) (2004). *Advances in spatial econometrics: methodology, tools, and applications*. New York, NY: Springer-Verlag. p. 359.

Armstrong, M. P., Rushton, G., and Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine* 18, pp. 497–525.

Arthur-Hartranft, S. T., Carlson, T. N., and Clarke, K. C. (2003). Satellite and ground-based microclimate and hydrologic analyses coupled with a regional urban growth model. *Remote Sensing of Environment* 86, pp. 385–400.

Atasoy, M., Palmquist, R. B., and Phaneuf, D. J. (2006). Estimating the effects of urban residential development on water. *Journal of Environmental Management* 79, pp. 399–408.

Atkinson, P. M., and Tate, N. J. (2000). Spatial scale problems and geostatistical solutions: a review. *Professional Geographer* 54, pp. 607–623.

Bae, C. H. C., et al. (2007). The exposure of disadvantaged populations in freeway air-pollution sheds: a case study of the Seattle and Portland regions. *Environment and Planning B – Planning & Design* 34, p. 154.

Berke, E. M., et al. (2007a). Protective association between neighborhood Walkability and depression in older men. *Journal of the American Geriatric Society* 55, pp. 526–533.

——. (2007b). Association of the built environment with physical activity and obesity in older persons. *American Journal of Public Health* 97, pp. 486–492.

Bolitzer, B., and Netusil, N. R. (2000). The impact of open spaces on property values in Portland, Oregon. *Journal of Environmental Management* 59, p. 185.

Brown, D. G. (2003). Land use and forest cover on private parcels in the Upper Midwest USA, 1970 to 1990. *Landscape Ecology* 18, pp. 777–790.

Brown, P. (2004). Devil is in the detail as OS maps out the future. *The Guardian* March 8 2004 5.

Bullen, N., Jones, K., and Duncan, C. (1997). Modelling complexity: analysing between-individual and between-place variation – a multilevel tutorial. *Environment and Planning A* 29, pp. 585–609.

Burra, T., et al. (2002). Conceptual and practical issues in the detection of local disease clusters: a study of mortality in Hamilton, Ontario. *Canadian Geographer* 46, pp. 160–171.

Carrion-Flores, C., and Irwin, E. G. (2004). Determinants of residential land-use conversion and sprawl at the rural-urban fringe. *American Journal of Agricultural Economics* 86, pp. 889–904.

Carruthers, J. I. (2002). The impacts of state growth management programmes: a comparative analysis. *Urban Studies* 39, pp. 1959–1982.

Cervero, R., and Duncan, M. (2007). Transit's value-added effects: light and commuter rail services and commercial land values. *Transportation Research Record* 1805, pp. 8–15.
Cho, W. K. T. (1998). Iff the assumption fits: a comment on the king ecological inference solution. *Political Analysis* 7, pp. 143–163.
Conway, T. M. (2005). Current and future patterns of land-use change in the coastal zone of New Jersey. *Environment and Planning B: Planning & Design* 32, p. 877.
Cowen, D. J., and Craig, W. J. (2003). A retrospective look at the need for a multipurpose cadastre. *Surveying and Land Information Science* 63, p. 205.
Craig, V. (1998). Kids mobility project. Minnesota, MN: Hennepin County Office of Planning and Development and Community Corrections..
Craig, W. J. (2005). White knights of spatial data infrastructure: the role and motivation of key individuals. *URISA Journal* 16, pp. 5–13.
Cunningham, C. R. (2006). House prices uncertainty, timing of development, and vacant land prices: evidence for real options in Seattle. *Journal of Urban Economics* 59, pp. 1–31.
Curry, M. (1997). The digital individual and the private realm. *Annals of the Association of American Geographers* 87, pp. 681–699.
Dearwent, S. M., Jacobs, R. R., and Halbert, J. B. (2001). Locational uncertainty in georeferencing public health datasets. *Journal of Exposure Analysis and Environmental Epidemiology* 11, pp. 329–334.
Doss, C. R., and Taff, S. J. (1996). The influence of wetland type and wetland proximity on residential property values. *Journal of Agricultural and Resource Economics* 21, p. 120.
Dungan, J. L., et al. (2002). A balanced view of scale in spatial statistical analysis. *Ecography* 25, pp. 626–640.
Ellis, C. D., Lee, S. W., and Kweon, B. S. (2006). Retail land use, neighborhood satisfaction and the urban forest: an investigation into the moderating and mediating effects of trees and shrubs. *Landscape and Urban Planning* 74, pp. 70–78.
Entwisle, B. (2007). Putting people into place. *Demography* 44, pp. 687–703.
FGDC (1998). *Content standard for digital geospatial metadata*. Washington, DC: Federal Geographic Data Committee.
——. (2003). *Cadastral data content standard for the national spatial data infrastructure*. Washington, DC: Federal Geographic Data Committee.
Fox, J., et al. (2003). *People and the environment: approaches for linking household and community surveys to remote sensing and GIS*. Boston, MA: Kluwer Academic Publishers.
Freeman, A. M. (2003). *The measurement of environmental and resource values: theory and methods*. Washington, DC: Resources for the Future.
Geoghegan, J., Wainger, L. A., and Bockstael, N. E. (1997). Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics* 23, pp. 251–264.
Georgiadou, Y., Rodriguez-Pabon, O., and Lance, K. T. (2006). Spatial data infrastructure (SDI) and E-governance: a quest for appropriate evaluation approaches *URISA Journal* 18, pp. 43–55.
Gewin, V. (2004). Mapping opportunities. *Nature* 427, pp. 376–377.
Haque, A. (2001). GIS, public service, and the issue of democratic governance. *Public Administration Review* 61, pp. 259–265.
Harvey, F., and Tulloch, D. L. (2006). Local government data sharing: evaluating the foundations of spatial data infrastructures. *International Journal of Geographical Information Science* 20, pp. 743–768.
Harvey, F., et al. (1999). Semantic interoperability: a central issue for sharing geographic information. *Annals of Regional Science* 33, pp. 213–232.
Haurin, D. R., and Brasington, D. (1996). School quality and real house prices: inter- and intrametropolitan effects. *Journal of Housing Economics* 5, pp. 351–368.
Hayward, S., and Mondale, T. (2000). The brawl over sprawl: is 'Smart Growth' really all that smart. *American Experiment Quarterly* 3, pp. 57–71.
Huxhold, W. E. (1991). *An introduction to urban geographic information systems*. New York, NY: Oxford University Press.
Huxhold, W. E., Fowler, E. M., and Parr, B. (2004). *ArcGIS and the digital city: a hands-on approach for local government*. Redlands, CA, ESRI.

Irwin, E. (2002). The effects of open space on residential property values. *Land Economics* 78, pp. 465–481.
Irwin, E. G., and Bockstael, N. (2004). Endogenous spatial externalities: empirical evidence and implications for the evolution of exurban residential land use patterns. In: Anselin, L., Florax, R. J. G. M. and Rey, S. J. (eds) *Advances in spatial econometrics: methodology, tools, and applications*. New York, NY: Springer-Verlag, p. 359.
Irwin, E. G., and Geoghegan, J. (2001). Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems, and Environment* 85, pp. 7–23.
Johnson, J. P. (1995). Case studies of dissemination policy in local government GIS agencies. *Computers, Environment and Urban Systems* 19, pp. 373–389.
Johnson, R. (2005). Minnesota Geospatial Data Collaborative, Minnesota–St. Paul Metropolitan Area. *URISA Journal* 17, pp. 41–45.
Kaufmann, J., and Steudler, D. (1998). *Cadastre 2014: a vision for a future cadastral system*. Copenhagen, Denmark: International Federation of Surveyors.
King, G. (1997). *A solution to the ecological inference problem: reconstructing individual behavior from aggregate data*. Princeton, NJ: Princeton University Press.
Klinkenberg, B. (2003). The true cost of spatial data in Canada. *The Canadian Geographer* 47, pp. 37–49.
Landis, J. D., et al. (2006). The future of infill housing in California: opportunities, potential, and feasibility. *Housing Policy Debate* 17, pp. 681–725.
Lee, C., Moudon, A. V., and Courbois, J. Y. P. (2006). Built environment and behavior: spatial sampling using parcel data. *Annals of Epidemiology* 16, pp. 387–394.
Liverman, D., et al. (1998). *People and pixels: linking remote sensing and social science*. Washington, DC: National Academy Press.
Lockyer, B. (2005). Opinions of the attorney general of California. *04–1105*. Sacramento, CA: Matthew Bender.
Longley, P. (2000). Spatial analysis in the new millennium. *Annals of the Association of American Geographers* 90, pp. 157–166.
Longley, P. A. (2003). Geographical information systems: developments in socio-economic data infrastructures. *Progress in Human Geography* 27, p. 114.
Lutzenhiser, M., and Netusil, N. R. (2001). The effect of open spaces on a home's sale price. *Contemporary Economic Policy* 19, p. 291.
Mansfield, C., et al. (2005). Shades of Green: measuring the value of urban forests in the housing market. *Journal of Forest Economics* 11, pp. 177–199.
Miranda, M. L., et al. (2007). Changes in blood lead levels associated with use of chloramines in water treatment systems. *Environmental Health Perspectives* 15, pp. 221–225.
Miranda, M. L., Dolinoy, D. C., and Overstreet, M. A. (2002). Mapping for prevention: GIS models for directing childhood lead poisoning prevention. *Environmental Health Perspectives* 110, pp. 947–953.
Moudon, A. V., et al. (2005). Cycling and the built environment, a US perspective. *Transportation Research Part D-Transport and Environment* 10, pp. 245–261.
——. (2007). Attributes of environments supporting walking. *American Journal of Health Promotion* 21, pp. 448–459.
Mugerauer, R. (2000). Qualitative GIS: to mediate, not dominate. In: Janelle, D. G. and Hodge, D. C. (eds) *Information, Place, and Cyberspace Issues in Accessibility*. Berlin, Germany: Springer Verlag, pp. 317–338.
Newburn, D. A., and Berck, P. (2006). Modeling suburban and rural-residential development beyond the urban fringe. *Land Economics* 82, pp. 481–499.
NRC (1980). *Need for a multipurpose cadastre*. Washington, DC: National Research Council of the National Academies.
——. (2003). *GIS for housing and urban development*. Washington, DC: National Academies Press.
——. (2007). *National land parcel data: a vision for the future*. Washington, DC: National Research Council of the National Academies.
Olivera, F., and DeFee, B. B. (2007). Urbanization and its effect on runoff in the Whiteoak Bayou watershed, Texas. *Journal of the American Water Resources Association* 43, pp. 170–182.

Orford, S. (2000). Modelling spatial structures in local housing market dynamics: a multilevel perspective. *Urban Stud* 37, pp. 1643–1671.
Peuquet, D. J. (1981a). An examination of techniques for reformatting digital cartographic data. Part I: the raster to vector process. *Cartographica* 18, pp. 34–48.
——. (1981b). An examination of techniques for reformatting digital cartographic data. Part II: the vector-to-raster process. *Cartographica* 18, pp. 21–33.
Polimeni, J. M. (2005). Simulating agricultural conversion to residential use in the Hudson River Valley: scenario analyses and case studies. *Agriculture and Human Values* 22, pp. 377–393.
Rajabifard, A., et al. (2007). Assessing the worldwide comparison of cadastral systems. *Land Use Policy* 24, pp. 275–288.
Ready, R. C., and Abdalla, C. W. (2005). The amenity and disamenity impacts of agriculture: estimates from a hedonic pricing model. *American Journal of Agricultural Economics* 87, p. 314.
Rhind, D. (1996). Economic, legal and public policy issues influencing the creation, accessibility and use of GIS databases. *Transactions in Geographical Information Systems* 1, pp. 3–11.
Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociology Review* 15, pp. 351–357.
Rodriguez, F., Morena, F., and Andrieu, H. (2005). Development of a distributed hydrological model based on urban databanks – production process of URBS. *Water Science & Technology* 52, pp. 241–248.
Roemmich, J. N., et al. (2006). Association of access to parks and recreational facilities with the physical activity of young children. *Preventive Medicine* 43, pp. 434–441.
——. (2007). The neighborhood and home environments: disparate relationship with physical activity and sedentary behaviors in youth. *Annals of Behavioral Medicine* 33, pp. 29–38.
Rogers, G. O., and DeFee, B. B., II (2005). Long-term impact of development on a watershed: early indicators of future problems. *Landscape and Urban Planning* 73, pp. 215–233.
Rull, R. A., and Ritz, B. (2003). Historical pesticide exposure in California using pesticide use reports and land-use surveys: an assessment of misclassification error and bias. *Environmental Health Perspectives* 111, pp. 1582–1589.
Sengupta, S., and Osgood, D. E. (2003). The value of remoteness: a hedonic estimation of ranchette prices. *Ecological Economics* 44, p. 91.
Shilling, F., and Girvetz, E. (2007). Physical and financial barriers to implementing a nature reserve network in the Sierra Nevada, California, USA. *Landscape and Urban Planning* 80, pp. 165–172.
Stage, D., and von Meyer, N. (2006a). *An assessment of best practices in seven state parcel management programs*. Washington, DC: FGDC Cadastral Data Subcommittee.
——. (2006b). *An assessment of parcel data in the United States – 2005 survey results*. Washington, DC: National Academies Press. [online]. Retrieved from http://www.nationalcad.org/data/documents/Cadastral_Inv_2005_v2.pdf
Steudler, D., Rajabifard, A., and Williamson, I. P. (2004). Evaluation of land administration systems. *Land Use Policy* 21, pp. 371–380.
Stone, B. (2004). Paving over paradise: how land use regulations promote residential imperviousness. *Landscape and Urban Planning* 69, pp. 101–113.
Stone, B., and Bullen, J. L. (2006). Urban form and watershed management: how zoning influences residential stormwater volumes. *Environment and Planning B – Planning & Design* 33, p. 21.
Straeger, M. P., and Rosenberger, R. S. (2007). Aggregating high-priority landscape areas to the parcel level: an easement implementation tool. *Journal of Environmental Management* 82, pp. 290–298.
Tajima, K. (2003). New estimates of the demand for urban green space: Implications for valuing the environmental benefits of Boston's big dig project. *Journal of Urban Affairs* 25, p. 641.
Tate, N. J., and Atkinson, P. M. (2001). *Modelling scale in geographical information science* Chichester, UK: John Wiley and Sons.
Theobald, D. M. (2001). Land-use dynamics beyond the American urban fringe. *Geographical Review* 91, pp. 544–565.
Treuhaft, S., and Kingsley, G. T. (2008). *Transforming community development with land information systems*. Cambridge, MA: Lincoln Institute of Land Policy.

Troy, A., and Wilson, M. A. (2006). Mapping ecosystem services: practical challenges and opportunities in linking GIS and value transfer. *Ecological Economics* 60, pp. 435–449.
Vance, C., and Iovanna, R. (2006). Analyzing spatial hierarchies in remotely sensed data: insights from a multilevel model of tropical deforestation. *Land Use Policy* 23, p. 226.
VanWey, L. K., et al. (2005). Confidentiality and spatially explicit data: concerns and challenges. *Proceedings of the National Academy of Sciences* 102, pp. 15337–15342.
Von Meyer, N. (2001). *GIS and land records: the parcel data model*. Redlands, CA: ESRI.
Waddell, P. (2000). A behavioral simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. *Environment and Planning B: Planning & Design* 27, pp. 247–263.
Waller, L. A., and Gotway, C. A. (2004). *Applied spatial statistics for public health data*. Hoboken, NJ: John Wiley & Sons Inc.
Williamson, I., Ting, L., and Grant, D. (1999). The evolving role of land administration in support of sustainable development. *The Australian Surveyor* 4, pp. 126–135.
Wong, D. (2004). The Modifiable Areal Unit Problem (MAUP). In: Warf, B. and Hansen, K. (eds.) *WorldMinds: geographical perspectives on 100 Problems*. Berlin, Germany: Springer, pp. 571–575.
Wu, J. J., Adams, R. M., and Plantinga, A. J. (2004). Amenities in an urban equilibrium model: residential development in Portland, Oregon. *Land Economics* 80, p. 19.
Zandbergen, P. A. (2007). Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *BMC Public Health* 7, p. 37.