*Department of City and Regional Planning*

*Departmental Papers (City and Regional Planning)*

*University of Pennsylvania* *Year* 2005

# Integrating and Distributing Administrative Data to Support Community Change

Amy E. Hillier [*] Dennis P. Culhane [†]

[*]University of Pennsylvania, ahillier@design.upenn.edu

[†]University of Pennsylvania, culhane@mail.med.upenn.edu

# The Handbook of COMMUNITY PRACTICE

Edited by
**Marie Weil**
*University of North Carolina, Chapel Hill*

Associate Editors:
**Michael Reisch**
*University of Michigan*
**Dorothy N. Gamble**
*University of North Carolina, Chapel Hill*
**Lorraine Gutiérrez**
*University of Michigan*
**Elizabeth A. Mulroy**
*University of Maryland*
**Ram A. Cnaan**
*University of Pennsylvania*

SAGE Publications
Thousand Oaks ▪ London ▪ New Delhi

CHAPTER 36

# Integrating and Distributing Administrative Data to Support Community Change

*Amy E. Hillier*

*Dennis P. Culhane*

An increasing number of groups are demanding access to information about the physical, economic, and social conditions in their communities. In addition to researchers, community development corporations, neighborhood associations, social service agencies, and municipal agencies all are becoming consumers of quantitative data. More powerful desktop computers, sophisticated data management software, enormous data storage devices, and expanding Internet access have increased exponentially the capacity of even small organizations to handle large amounts of data. However, none of these advances guarantees that organization staff will be able to make sense of these mountains of data. Perhaps more important, the advances do not enhance access to the most useful types of data that frequently reside behind municipal agency firewalls and layers of bureaucratic red tape. The growing demand for small-area administrative data in useful formats makes essential the development of new tools to support a wide range of community change efforts.

## Need for Small-Area Data

Data about the well-being of demographic groups, families, institutions, the built environment, the natural environment, and businesses play an increasingly important role in a number of efforts to improve communities. Analysis at the individual person or property level is often impractical and perhaps unethical because of limited access owing to concerns about confidentiality and the amount of information contained in individual-level data sets. Data aggregated by small geographic areas, such as census tracts, neighborhoods, and service districts, avoid problems with confidentiality and allow for easier recognition of patterns across space and time. Access to small-area data can facilitate needs assessments, program planning, site selection, and resource allocation by public or private agencies. Small-area data can also support long-term community planning, serving as the basis for forecasts of demographic and housing changes as well as models of the desired impact that various programs might have. The increasing emphasis on measurable outcomes by funders, including private foundations and the federal government, has also increased the need for trend data to evaluate the impact of programs. The ability of communities to monitor the impact of government and private actions, including their ability to mount Community Reinvestment Act challenges, also depends on their access to appropriate small-area data.

The demand for small-area data has increased over the past century along with the number of small geographic units available to serve as units of analysis. In his classic small-area study of housing and economic conditions, *The Philadelphia Negro,* W. E. B. DuBois used the political ward as his focus (1899). Responding to the need for new geographic units that were not politically based and were less subject to boundary changes, the Census Bureau introduced census tracts (originally called districts) in 1910. Concerned that cities did not recognize the value of these new geographic tabulations, the American Statistical Association established a Committee on Census Enumeration Areas in 1931 to promote census tracts and identify uses for tract-level data (U.S. Department of Commerce, 1994). President Hoover's Committee on Social Trends relied on a combination of ward-, municipal-, and county-level data for *Recent Social Trends,* the comprehensive account of social conditions it published in 1933 (President's Research Committee on Social Trends, 1933). On the other hand, the Works Progress Administration Real Property Surveys, conducted later that decade, used census tracts and blocks to report extensive information about households and housing conditions. The U.S. Bureau of the Census first provided data at the block level in 1940, followed by the block group in 1970.

The Social Indicators movement, encouraging the use of time-series data to monitor and effect social change, has increased considerably the demand for small-area data as it has gained momentum over the past 30 years (Bauer, 1966; Sawicki & Flynn, 1996). The Urban Institute's National Neighborhood Indicators Project has been supporting indicator projects in six cities—Atlanta, Boston, Cleveland, Denver, Oakland, and Providence—since the late 1980s (Kingsley, 1998). Additional projects in Jacksonville, Seattle, Pasadena, Milwaukee, and a host of other cities also have served to integrate and interpret small-area data to support community change efforts. Census data play an important role in these types of projects. However, the limitations imposed by the decennial nature of census data, lack of correspondence among local service areas and planning districts with the U.S. Bureau of the Census's

standard geographic units, and the limited range of data elements made available have led researchers and community groups to look elsewhere.[1]

Municipal governments are perhaps the largest potential source of small-area data. City agencies are involved either directly or indirectly as conduits of city, state, and federal funding in nearly all areas of service delivery and community development, from public assistance, child welfare, education, and health care to housing inspections, rehabilitation, and demolition. With some notable exceptions, community organizations need to cajole or pay individual municipal agencies for administrative data or they have to appeal to their legal right to such information through Freedom of Information Act requests. In some cases, community organizations can acquire information about individual properties or areas but have much greater difficulty acquiring complete data sets that allow for citywide analysis.

## Overview of Web-Based Information Systems

Web-based information systems that integrate and distribute administrative data aim to help fill the need left by the decennial U.S. Census. These systems take advantage of many technological advances in database administration, programming, and computer mapping. But they also depend on strong collaborations among their developers, data-providing agencies, and system users as much as on technical expertise. Although there are few limits to the forms these systems can take, there are a number of principles that distinguish them. Most important, they distribute data, making hard-to-acquire information accessible to a wide array of user groups. Although mapping may or may not play a large role in data displays, common geographic references—either street addresses or small-area units—provide the basis for data integration. These systems are both replicable and customizable. Much of the computer programming that creates the applications that serve up the data—including functionality to integrate, summarize, chart, map, and print the data—is transferable across applications that have different themes or serve different geographic areas. At the same time, these applications can adapt to specific data needs and data availability in individual municipalities. These systems are constantly maintained and updated. Data values are updated regularly to reflect continuous changes in local conditions. These systems are user-friendly, requiring minimal training for users who may have very limited computer experience and maximizing the ability of users to define their own queries and interpret the results accurately. Finally, the developers of these systems are accountable to funders, data providers, and system users.

### Examples of Web-Based Information Systems and Partnerships for Information Collection and Distribution

#### *Chicago's Neighborhood Early Warning System and Neighborhood Knowledge Los Angeles*

Chicago's Neighborhood Early Warning System (NEWS) provides one of the earliest examples of an information system created to prevent housing abandonment

by identifying at-risk properties. Started in 1984 when Harold Washington was mayor, NEWS originally relied on floppy disks to distribute data. Now on the Web and updated regularly, NEWS integrates property-level and area data on housing code violations, housing court cases, fires, tax assessments, and general housing characteristics for Chicago and Cook County (Center for Neighborhood Technology, 2003). Modeled on NEWS, Neighborhood Knowledge Los Angeles (NKLA) combines similar property-level and area housing data with U.S. Census data. In addition to generating tables based on inquiries about specific properties and areas, NKLA users can create maps and conduct queries to identify both the attributes of particular places as well as places with specified attributes (Community Information Technology Center, 2003).

### *The Cartographic Modeling Laboratory (CML) and Partnerships*

The CML, a joint venture of the University of Pennsylvania School of Social Work and Graduate School of Fine Arts, has developed several similar applications that distribute a wide range of administrative data to community and municipal government partners in Philadelphia. The Neighborhood Information System (NIS) provides information about properties and the built environment, whereas the Services Utilization Monitoring System is focused on human services data relating to children.

### *Philadelphia Neighborhood Information System (NIS)*

The Philadelphia NIS (Cartographic Modeling Laboratory [CML], 2003) integrates and distributes housing data from eight different municipal agencies along with 1990 and 2000 Census data through two different Web-based applications: ParcelBase and NeighborhoodBase. Agencies that provide data are the Office of Housing and Community Development, City Planning Commission, Department of Revenue, Department of Licenses and Inspections, Board of Revision of Taxes, Water Department, Philadelphia Gas Works, and U.S. Postal Service. Funding has been provided by the William Penn Foundation, the Pew Charitable Trusts, the University of Pennsylvania, and the city of Philadelphia.

The applications use Microsoft Visual Basic with ESRI's (Environmental Systems Research Institute's) ArcIMS (Arc Internet Map Server) to serve up various data displays in response to user requests. Data about property sales, ownership, housing vacancy, utility terminations, tax arrearages, housing code violations, and demolitions are featured. Data elements to be included were determined through discussions between the CML and data-providing agencies as well as through focus groups held with community organizations. The data-providing agencies are asked to provide data updates every three months.[2]

ParcelBase provides property-level data to registered users who consist of staff at city agencies and community organizations authorized to have an account by the city's Office of Housing and Community Development. Users identify a property by typing an address or property identification number, consisting of a five-digit street code and the house number, or by highlighting a property on a parcel map. A small screen pops up when a parcel is selected, displaying characteristics

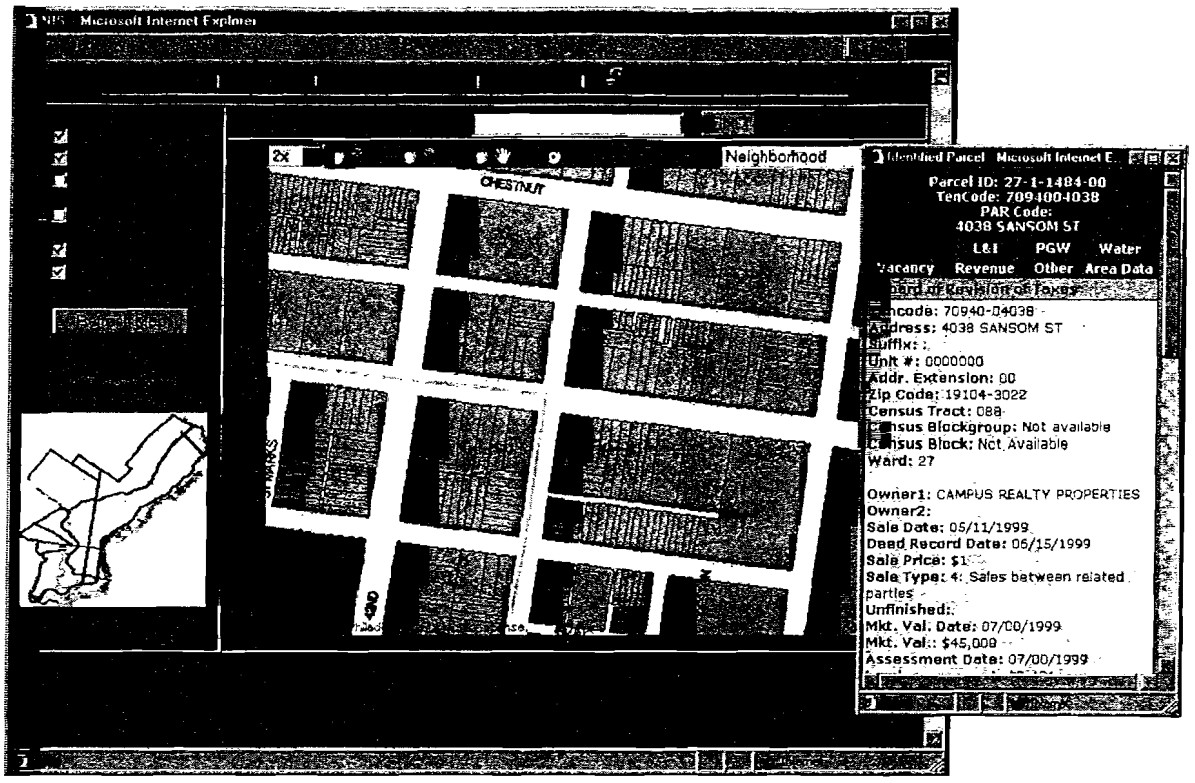


**Figure 36.1** ParcelBase Application

*Source:* http://cml.upenn.edu/parcelbase/

*Note:* The ParcelBase application integrates property-level housing data from eight different municipal agencies in Philadelphia. Property characteristics are grouped according to their source agency using eight different tabs on a screen that pops up when a parcel is selected.

of the properties grouped according to their source agency using eight different tabs (see Figure 36.1). For properties identified as vacant through a foot survey conducted in certain parts of the city, a digital photograph of the property is also displayed. The query function allows users to create a list or map of properties that meet the criteria they identify. Users can generate and print reports that include a small map of the property and the characteristics from each of the agencies. ParcelBase went live in June 2000. As of this writing, there are more than 500 registered NIS users from more than 150 community organizations and 50 city agencies.

The NeighborhoodBase provides public access to the same data, aggregated by six different geographic units: Census tracts, Census block groups, councilmanic districts, neighborhoods, zip codes, and elementary school feeder areas. Data are aggregated using a georeference file that identifies the tract, block group, councilmanic district, neighborhood, zip code, and school area in which each address falls. The application then serves up the data from the preaggregated tables. Users can generate summary statistics, queries, tables, charts, or maps (see Figure 36.2). Through of series of drop-down menus, users then choose a general content area—housing characteristics, land use, property ownership, property taxes and revenue, real estate sales, utility information, vacancy and abandonment, and U.S. Census data—specific data element, year, and geographic unit. Neighborhood

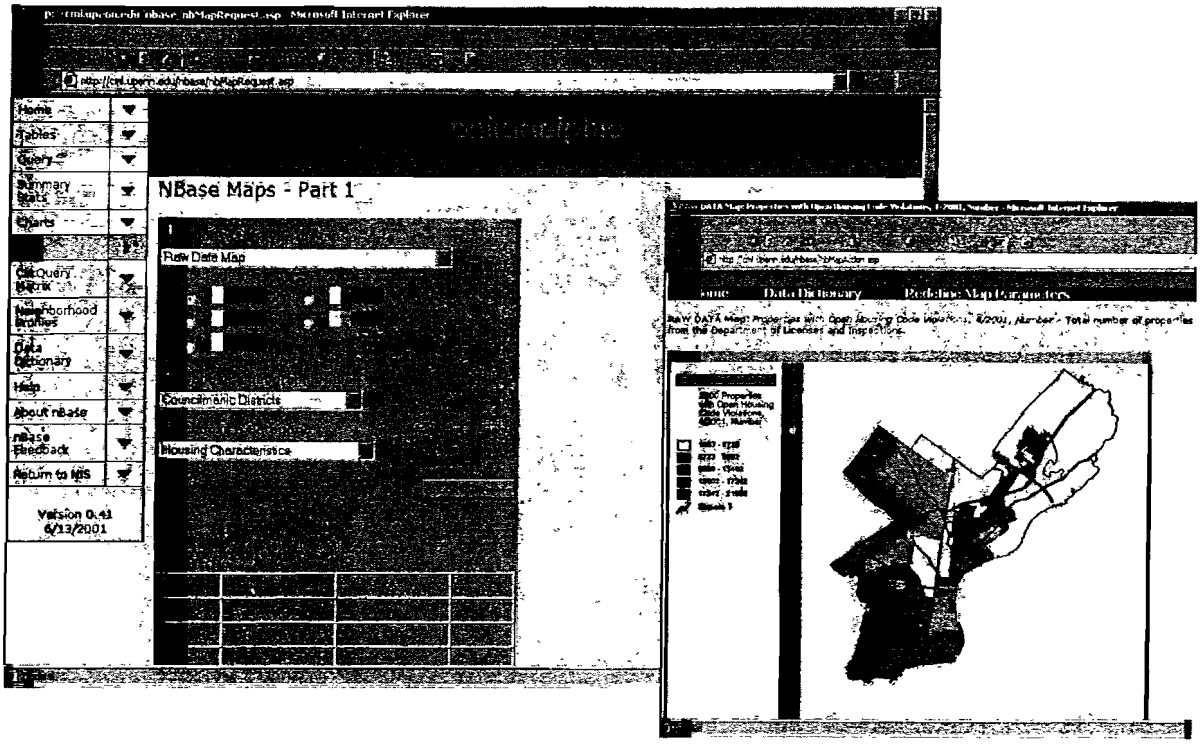


**Figure 36.2** NeighborhoodBase Application

*Source:* http://cml.upenn.edu.nbase/

profiles provide maps identifying the location of each neighborhood (including boundaries, schools, libraries and recreation centers), neighborhood trivia, and summary data.[3]

*NeighborhoodBase Application.* The NeighborhoodBase allows users to create tables, charts, and maps with the same data included in ParcelBase aggregated at one of eight geographic units.

## *Services Utilization Monitoring System (SUMS)*

SUMS grew out of discussions with the agencies that serve Philadelphia's children and families.[4] The participating agencies, including the school district and the city's public health and child welfare departments, were interested in sharing data to improve planning and interagency coordination, but they faced data-sharing restrictions due to confidentiality regulations. None of the agencies are permitted to release individual-level data to each other on a regular or periodic basis, but they are willing to share their data if the data are aggregated by Census blocks or larger geographic units.

The aggregation occurs through one of two methods. Agencies can install a geocoding and aggregation software created by the CML for this purpose, or they can enlist the CML to do the aggregation onsite and under their administrative control. The SUMS application is restricted to registered users, as is ParcelBase, but

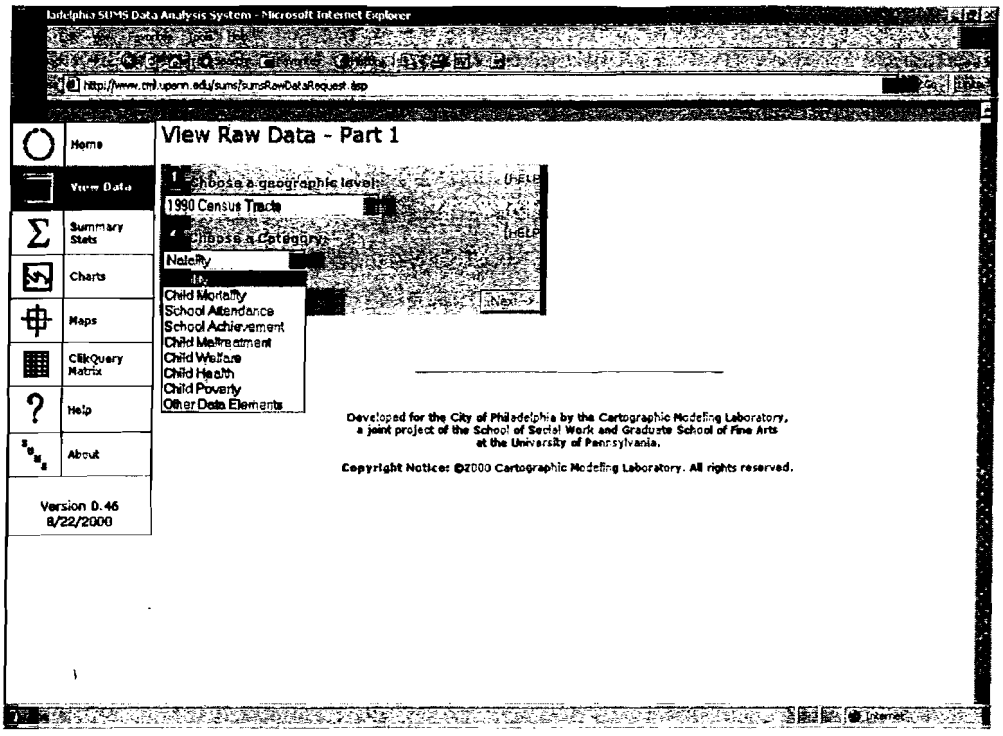


**Figure 36.3** Services Utilization Monitoring System (SUMS) Application

*Source:* http://apollo.gsfa.upenn.edu/project_areas/philadelphia_sums.htm

*Note:* Users identify a geographic area, data category, then a specific data element to create a table, chart, or map in SUMS.

otherwise it is very similar to NeighborhoodBase. Users can opt to generate summary statistics, tables, charts, or maps. They then choose from one of eight categories—natality, child mortality, school achievement, school attendance, child maltreatment, child welfare, child health, and child poverty—through which the more than 150 specific data elements are organized (see Figure 36.3). Finally, users select a year and geographic unit: 1990 Census tracts, 1990 block groups, elementary school feeder areas, school district clusters, councilmanic districts, or public health districts.

## Challenges in Developing and Maintaining Applications

Development of information systems like these involves a number of challenges. These range from building relationships with data-providing agencies, identifying appropriate data elements, and obtaining clear data documentation and regular updates, to assessing data quality, determining data access policies, developing and maintaining user-friendly applications, remaining accountable to supporters and users, developing long-range funding plans, and evaluating the processes and outcomes associated with the projects.

Establishing strong working relationships with data providers requires awareness of interagency dynamics. All or some of the municipal agencies that maintain housing data or information about children, for example, may not have a

history of working together. From the beginning, the NIS had a strong advocate in the director of the Office of Housing and Community Development, who wrote letters of introduction encouraging various city agencies to participate in the project. Face-to-face meetings—at both the beginning and later stages of the project—provide critical opportunities for communicating the purpose of the project and remaining accountable to the data providers. Changes in local administrations and staff at participating agencies complicate the relationship-building process. The SUMS project was developed with the support of one mayor and his administration, but his replacement chose not to continue to fund the project. To be successful, information systems require long-term support from city officials, regardless of whether the city is involved in funding them.

Some of the challenges involved in building trust and consensus among city agencies may be avoided if an agency within the city serves as the developer. However, there may not be an agency with the time and staff appropriate for developing and hosting an integrated information system. Municipal firewalls may also make distribution of data outside city networks extremely difficult. Furthermore, city agencies may have little incentive to share data with community organizations. Universities or other private entities represent alternative host sites. Universities may be particularly interested in such a role, given that researchers have a direct interest in access to small-area data for research purposes. Universities may also have relatively greater flexibility for hiring staff and coordinating development contracts.

Identifying the data to be included also requires considerable thought. Agency staff members know the most about their own data and they may also be able to indicate what data maintained by other agencies would be helpful to them. Community organizations must also be involved in the process of deciding what is included in the application, either through focus groups or through some sort of advisory group. Face-to-face meetings and presentations to data providers and groups of system users also serve to keep developers accountable. Although initial planning should be as thorough and inclusive as possible, the process of updating and upgrading these information systems must also be flexible enough to incorporate suggestions for additions and improvements on an ongoing basis.

Even when there are strong relationships among developers and top officials at data-providing agencies, actually acquiring data regularly may still pose significant challenges. Relationships must be cultivated and maintained with the staff responsible for preparing the data in addition to those who authorize the data sharing. For both the NIS and SUMS, a schedule for updates was established up front and letters requesting data updates are sent several weeks prior to the expected transfer of data. Ideally, data updates would occur electronically and involve files in a standard format. But developers need to accept data in the medium and format available to data providers. Developers must also adapt to changes over time within the databases maintained by data providers, accepting that updates may not always include identical data.

Determining what level of data can be shared with which groups of users also involves some negotiation. For housing data, it may be appropriate to share address-level information with a wide array of groups. However, sharing very small-area aggregate human services data can pose threats to privacy. Care should be taken to

ensure that data aggregated by small-areas such as city blocks do not allow for the identification of individuals. Once it is determined which user groups should have access to what level of data, data agreements should be formalized through letters of authorization from the data-providing agencies. The goal of these systems is to expand access to administrative data, and whenever possible, these systems should be available to the public. However, such access should not jeopardize the privacy of those who participate in city services—populations which by definition are generally vulnerable. Nor should it jeopardize relationships with data providers. Discussions of which administrative data is "public" and what that means in legal and ethical terms are an important by-product of the development of these information systems.

The goal of these Web-based applications is to increase access to information that can support community change efforts, not just to distribute raw administrative data. Therefore, users must be confident in the quality of the data. Some amount of random error is inevitable given the size of the data sets acquired; however, care should be taken to identify and resolve systematic errors. Some administrative data may be maintained for reporting purposes, but more likely they are part of a billing or service delivery system, further complicating the translation into useful formats for users outside the agency. Ideally, all agencies would use the same unique property identifier to facilitate data integration, but this may not be the case. The development team should include a database administrator who reviews all data for inconsistencies. Distributing administrative data also introduces an additional layer of quality control, and these systems should make it simple for users to report possible problems.

The applications distributing the data must also be user-friendly. Users should not be overwhelmed by the amount of data or the functionality, but the applications should be able to display the data in a number of formats including simple charts, tables, and maps. Online help systems that explain the different options are essential. Developers have a responsibility to encourage careful use of data, and users should be guided in their decisions to normalize and display data to prevent inappropriate choices. Data dictionaries should also be accessible, explaining the meaning of each of the data elements included. Such dictionaries rely on the establishment of clear metadata systems at the outset of development that indicate the source of various data sets, the date of acquisition, restrictions on access, and explanations of the variable names and values.

Given the emphasis they place on geography, these applications should allow users to choose from a number of geographic units. Just as with the identification of appropriate data elements, user groups should be consulted in deciding what geographic units should be included. These may range from standard census geographies, to local political subdivisions, to various service areas, planning districts, and neighborhood definitions. This may require time-consuming digitizing of paper maps, but the value of the small-area analysis depends on how meaningful the small-areas are to different users.

To ensure the sustainability of these systems, long-term funding plans need to be developed at the outset. Although private foundations may be appropriate sources of funding for initial development, user groups or city governments probably need to play a role in covering costs in the long run. This might be

accomplished through a single contract with a municipal agency or a user fee based on the number of users or amount of time particular organizations require. However, city agencies and community organizations generally find funding for such services difficult given their limited budgets. Chicago's NEWS and NKLA have both relied on multiple sources of funding. The city of Chicago supported the NEWS project initially, and data updates are currently supported by a private foundation. NKLA has been funded through the National Telecommunication Information Administration, Fannie Mae Foundation, and the Los Angeles Housing Department, as well as the Microsoft Corporation. Funding for the first three years of the NIS came from two large private foundations, the Pew Charitable Trusts and the William Penn Foundation, as well as the University of Pennsylvania. Since then, the city of Philadelphia has joined the list of funders.

These systems also need to be evaluated on an ongoing basis using both process and outcome measures. Process evaluations might focus on the amount and extent of participation in developing the application. How cooperative were city agencies? Were they willing to dedicate staff time to identify appropriate data elements and transfer data regularly? How accountable was the developer to data providers and groups of users? Do potential users have an opportunity to influence the content, functionality, and levels of access to the information? These types of questions might be addressed through an outside evaluation or on an ongoing basis through discussions and online feedback forms. Measuring outcomes poses a greater challenge. In what ways did the application support community change efforts? How did community groups use the information generated? To what extent did the information save organizations time and money that otherwise would have been spent collecting data? To what extent did community and city agency staff develop skills for interpreting quantitative data? Initial development should spell out goals and how and by whom progress will be measured.

## New Opportunities

Despite the significant challenges involved in their development, these information systems have tremendous potential for supporting community change efforts by efficiently distributing previously inaccessible data to large numbers of community change agents. Beyond merely sharing data, these systems can provide simple yet powerful tools for data analysis that help even those who have limited experience with quantitative data to identify and interpret the meaning of patterns in demographics, physical, economic, environmental, and health conditions, and service delivery. Geographic information system and data analysis software can be expensive and complicated. Online applications such as the Philadelphia NIS can deliver the necessary functionality without overwhelming and confusing users. Well-designed applications can help users generate and interpret summary statistics, such as frequencies and measures of central tendency. The inclusion of mapping functionality also encourages users to think spatially and recognize spatial clustering, sparseness, and associations. More sophisticated applications may also be able to introduce intermediate users to basic inferential statistics such as

chi-square tests and linear regression. Universities are the appropriate teachers of basic quantitative data analysis skills, either through train-the-trainer programs with umbrella associations that provide technical assistance to community groups or through direct training programs and student internships.

There are no limits to the types of applications that might be developed beyond limits to funding and access to raw data. Web-based applications could be developed around a public safety theme and share information about nuisances, property crimes, and violent crimes, along with parking and housing violations. In addition to supporting the data needs of police and other public safety officials, these systems could support the work of police advisory groups, town watches, and other community organizations. In addition to providing information, they might provide opportunities for filing complaints, paying tickets, or communicating with public officials through virtual chat rooms. A public health application could integrate data about births and deaths, lead paint contaminations, immunizations, dangerous housing, communicable diseases, health insurance coverage, and access to health care facilities for all age groups. Local commerce departments, chambers of commerce, or community development corporations might sponsor an application focused on economic development. This could integrate information about types of businesses, business revenues, employment levels, commuting patterns, and day-care facilities, as well as locations of contiguous open land and brownfields for new development.[5]

These information systems need not be limited to administrative data. Community organizations can also contribute their own data based on foot surveys or questionnaires. With support from the CML and the Philadelphia Association of Community Development Corporations, several community organizations have conducted foot surveys of housing in their service areas. Surveyors used handheld computer devices programmed with survey questions and digital cameras to capture information about housing conditions not recorded by the municipal agencies. To facilitate integration into the system, NIS staff gave surveyors guidelines for the resolution and property identification for the digital photos. Community organizations can also generate and contribute more qualitative data to provide context for the quantitative administrative data, including photographs, interviews, and narratives.

At their best, these information systems provide opportunities for widespread understanding of current conditions that can be used to build consensus toward needed change. They empower individuals and organizations that might otherwise expend enormous effort to acquire small amounts of information, inviting them into broad discussions about the communities in which they live and work. They also provide objective standards for planning and evaluation, encouraging agents of community change to base decisions on empirical data rather than conjecture, anecdotal information, or political considerations. They also hold the potential to make municipal government more accountable to its citizens by sharing the information about the services it provides. This democratic ideal may not be realized in all existing distributed information systems, because developers face limits in funding and the willingness of city agencies to share information. However, this ideal represents the standard toward which developers should aim.

## Notes

1. If the U.S. Bureau of the Census moves forward with more frequent administration of the long form through the American Community Survey initiative, Census data could prove more valuable, with annual information about population characteristics at the tract level. However, because it will depend on sampling, concerns about confidentiality will prevent small-area aggregation, such as block groups and Census blocks.

2. The one exception is the U.S. Post Office, which maintains information about vacant properties where mail is not delivered and provides data updates annually.

3. Neighborhood Information System (NIS) developers recognize that there are numerous ways to identify Philadelphia's neighborhoods. The NIS uses a map with 69 neighborhoods that were determined by the Temple University Social Science Data Library based on information gathered from the Philadelphia Police Department, the Philadelphia Inquirer, and historical research. Neighborhood boundaries are coterminus with 1990 Census tract boundaries.

4. The Cartographic Modeling Laboratory has developed a similar application for Palm Beach County, Florida, and is currently developing a system for Columbus, Ohio.

5. The Cartographic Modeling Laboratory developed a brownfields application that integrated vacant land information for part of North Philadelphia into a stand-alone mapping application for Philadelphia's Department of Commerce and the Pennsylvania Environmental Council.

## References

Bauer, R. A. (Ed.). (1966). *Social indicators.* Cambridge, MA: MIT Press.

Cartographic Modeling Laboratory. (2003). *Philadelphia Neighborhood Information System.* Retrieved on April 29, 2003, from http://www.cml.upenn.edu/nis.

Center for Neighborhood Technology. (2003). *Chicago Neighborhood Early Warning System.* Retrieved on April 29, 2003, from http://www.newschicago.org.

Community Information Technology Center. (2003). *Neighborhood Knowledge Los Angeles.* Retrieved on April 29, 2003, from http://nkla.sppsr.ucla.edu.

DuBois, W.E.B. (1899). *The Philadelphia Negro.* Philadelphia: University of Pennsylvania.

Kingsley, G. T. (1998). *Neighborhood indicators: Taking advantage of the new potential.* Washington, DC: Urban Institute.

President's Research Committee on Social Trends. (1933). *Recent social trends.* New York: McGraw-Hill.

Sawicki, D. S., & Flynn, P. (1996). Neighborhood indicators: a review of the literature and assessment of conceptual and methodological issues. *Journal of the American Planning Association, 62*(2), 165–183.

U.S. Department of Commerce, Bureau of the Census. (1994). *Geographic areas reference manual* (10:2–10:3). Retrieved April 29, 2003, from http://www.census.gov/geo/www/garm.html.