

Data Warehouse

Introduction and Background

A data warehouse will be the foundation of Detroit's Neighborhood Indicators System. Although many actors need quality data on neighborhoods, no one place currently provides this essential information. The data warehouse will solve that problem by making relevant data available in a central location, accessible through different applications. (Two initial applications are described here. In the future, there potentially may be many more). The information then can be used in a variety of ways by a variety of users.

Creating and maintaining a hub for neighborhood indicators data is a challenging endeavor. This section will discuss the challenges and present some ideas for how a data warehouse might be implemented in Detroit.

Methods and Results

Numerous neighborhood indicator systems exist in the United States and each system has a data warehouse platform. Each data warehouse serves the goals of its respective programs, and functions vary by system. In general these functions include:

- **Collection and Maintenance:** All systems serve some data collection and maintenance functions. They collect, organize and clean data from various sources. The National Neighborhood Indicators Partnership (NNIP), a collaborative effort by the Urban Institute to foster usage of neighborhood-level information promotes this function as a key aspect of neighborhood indicators systems.⁷ Also, compiling metadata — that is, information about the origin, date, source and type of data⁸ — on all data in the system is necessary to maintain a data warehouse.
- **Distribution:** All systems provide data to organizations and individuals in response to requests for that data via a web site. How this is done depends on the system and its goals and users. Table 1 below outlines some major methods, characteristics, and users.
- **Expansion:** Some data warehouses are designed for community-focused data uploading. This allows for

⁷ National Neighborhood Indicators Partnership. <http://www.urban.org/nnip/>. Accessed February 12, 2004.

⁸ For more information about metadata, please see p. 21.

Table 1
Data Warehouse Distribution Methods, Characteristics and Primary Users

Distribution Method	Characteristics	Primary User(s)
Pre-designed area reports	Reports with fixed areas (e.g. census tract) and data contents.	Individuals or organizations with fewer technical skills or in need of commonly used data (e.g., Census).
Data files	Spreadsheets or databases (in Microsoft Excel or Access, for instance).	Lower skill users in need of a basic database to begin to understand an area as well as high skill users looking to manipulate data and/or use it in geographic information systems (GIS) for analysis and to generate maps.
Custom built area reports	Reports for which the user can define the data included for specific areas.	Higher-skill community organizations or others with particular questions or reporting needs.
Maps (custom built)	Data is presented spatially, allowing an easier visual interpretation.	Higher-skill community organizations or others with particular questions or reporting needs.
Maps (pre-designed)	Data is presented spatially, allowing an easier visual interpretation.	Individuals or organizations with fewer technical skills or in need of commonly used maps (e.g., Census).

flexibility, and lets users with the capacity to collect data and internet access to contribute to the database. The West Oakland Environmental Indicators Project is an example of this functionality.

These components are not mutually exclusive, but no one system has all of these qualities. The strongest data warehouse system will include the best qualities from all of the examples.

Recommendations

A data warehouse should be established as the foundation for two major user-friendly systems – Early Detection System and Community Empowerment System. For these purposes, the establishment of a data warehouse presents challenges and requires some decisions. These issues and challenges involve data control and maintenance, data selection, and data acquisition. Each of these issues and their attendant considerations are discussed in turn to identify recommended directions for the Detroit Neighborhood Indicator System.

Data Control and Maintenance

The control and maintenance of the data warehouse has important implications for the neighborhood indicators project. The Detroit NIS will need an organization to function as a system administrator to host, collect, clean, and maintain the

data. To decide what organization should serve as the data host in Detroit, one must consider what makes for a strong intermediary. Characteristics of a strong candidate include:

- **Ability to provide or capture initial development and ongoing funding:** Many NIS are able to get off the ground through funding from grant-giving organizations like the MacArthur Foundation, Fannie Mae Foundation, Local Initiatives Support Corporation (LISC), The Enterprise Foundation and others. These groups generally do not provide funding for ongoing system maintenance. To be successful in the long term, the system administrator must either be able to fund maintenance costs or must be adept at capturing new grants.⁹
- **Staff availability:** Data warehouses must be constantly maintained. Thus, a strong intermediary must have the staff capacity to develop and maintain a web-served database. Based on research on other systems, this plan calls for a full-time database developer whose sole commitment is to the neighborhood indicator system.¹⁰
- **Technology resource availability:** To reduce development costs (both time and money) an intermediary would ideally already have sizable network and hosting capabilities which NIS servers¹¹ could have access to.
- **Ability to build trust and facilitate data sharing:** This is probably the most important trait of a good data intermediary. To build a strong neighborhood indicator system, staff working for the intermediary organization will build and maintain partnerships with city agencies, community organizations, and other data providers.

⁹ Funding such an organization will certainly be a challenge. Many programs use a variety of government and foundation grants. Government grants include the U.S. Department of Commerce Technology Opportunity Program, which is distributing \$12.9 million in fiscal year 2004. NNIP also discusses charging fees for data as a way to pay for the program. Some data intermediaries have a moderate annual membership fee that entitles the member to any amount of data.

¹⁰ Sanders, Greg. Northeastern Illinois Planning Commission. Personal Interview. February 25, 2004.

¹¹ A server is defined as a computer where data or other files are stored and can be requested by other computers on a network, such as the Internet. It allows multiple computer users to access files from the same shared computer space.

Different organizations function as data intermediaries for neighborhood indicator systems across the country. They are discussed below with respect to how well they fit the above criteria. The major groups are:

1. A municipal agency

Pro:

- A Detroit city governmental agency (i.e., P&DD) has qualified staff that can handle data collection and management, as well as data distribution. The agency also may have adequate computer resources that can handle large amounts of data.

Con:

- The city might not have sufficient staff time or

funding available to devote to the NIS.

- Some organizations might hesitate to give sensitive data such as crime data and welfare statistics to a city agency, in part because of confidentiality concerns.
- A sophisticated web-based approach also might require specialized staff training, new staff or consulting arrangements that would again infringe upon current staff time and funding allocations.
- Detroit agencies have been reluctant to share data and thus might have a credibility problem when asking other agencies and organizations to share their information.

Precedent: Miami-Dade County houses a data warehouse and distributes data upon request from community groups.¹²

2. University-based group

Pro:

- University-based groups usually have the capacity to handle data, with knowledgeable staff and computer resources to collect, store, manage and distribute information.

Con:

- A university partner would need agencies' trust.
- A university would probably rely on grant funding, and the program could face elimination if the university's commitments with individual partners or funding changed.

Precedent: The University of Pennsylvania heads a collaboration that handles data for a Philadelphia neighborhood indicators project.¹³

3. Non-profit organization

Many cities establish non-profit organizations to handle the data warehouse function.

Pro:

- Non-profit organizations that clearly communicate their mission and create confidentiality safeguards might have an easier time collecting data from various agencies than a governmental agency or university-based group.

Con:

- Establishing a new non-profit organization (rather than a sub-group from an existing entity) would require more effort and money and will necessarily involve job training, new staff hires and other administrative issues.

Precedents: A non-profit handles the data warehouse in Baltimore¹⁴ and Washington, D.C.¹⁵

¹² Kerr, Oliver. Miami-Dade Department of Planning and Zoning. Personal Interview. January 29, 2004.

¹³ Philadelphia Neighborhood Information System. <http://cml.upenn.edu/nbase/>. Accessed January 27, 2004.

¹⁴ Baltimore Neighborhood Indicators Alliance. *Research, Reports and Projects Produced by the Alliance*. http://www.bnia.org/results/research_reports.html. Accessed January 27, 2004.

¹⁵ D.C. Agenda Neighborhood Information Service. <http://www.dcagenda.org/nis/index.html>. Accessed January 30, 2004.

4. Regional governmental agencies

Some data warehouses are hosted by regional planning organizations. These organizations are usually concerned with neighborhood health across an entire region.

Pro:

- Information sharing, resource coordination and collaborative planning fall under the mission of most regional planning organizations. In addition, because these organizations do not speak for an individual jurisdiction, they might have an easier time than city governmental departments collecting data from various agencies.
- Generally, interested parties have less fear of regional planning organizations due to their lack of power at the local level. As system administrator this is an asset, as data gathering by coercion works far worse than offering incentives to recalcitrant data sources.

Con:

- In order to engage a regional planning organization, a neighborhood indicator system must take a regional view rather than a Detroit-centric view. P&DD and SDBA have defined the mission of this plan as Detroit-centric.

Precedents: Northeast Illinois Planning Commission's Greater Chicago Area Housing and Community Development Website.¹⁶

5. Foundations

Some data warehouses are hosted by foundations and other philanthropic groups.

Pro:

- Foundations likely will have the capacity to hire staff and commit capital resources, and may even already retain staff for the purpose of analyzing data.
- Foundations may have more success building coalitions between city agencies and other data providers.

Con:

- The mission of the foundation would likely dictate the direction and priority of the neighborhood indicator system, and would not necessarily fit all user organization's needs.

Precedents: The Piton Foundation manages Denver's neighborhood indicator system.¹⁷ The Boston Foundation coordinates a system in Boston¹⁸ in partnership with the City of Boston/Boston Redevelopment Authority and the Metropolitan Area Planning Council.

¹⁶ Greater Chicago Housing and Community Development Website. <http://www.chicagoareahousing.org/HousingHomePage.asp>. Accessed February 26, 2004.

¹⁷ Piton Foundation, The. <http://www.piton.org/>. Accessed January 24, 2004.

¹⁸ Boston Foundation, The. *Boston Indicators Report 2002*. <http://www.tbf.org/indicators/summary/index.asp>. Accessed January 24, 2004.

If the NIS moves forward, several potential partners would likely come forward to host the data warehouse. Based on conversations with hosts and our partners in Detroit, either a university department or a non-profit best fits the profile of a strong system administrator. In fact, most neighborhood indicators systems around the United States house the data warehouse in either a university or non-profit setting. Depending on the individual partner, either type could provide the fund raising, technical, and bridge-building skills needed for the success of a data warehouse. A phased approach should be considered where:

1. P&DD assembles as much city data as possible into a data warehouse and sorts those data into appropriate categories like housing, demography, economy, transportation to allow flexibility and easy data access.
2. The Detroit P&DD and community-based organizations cooperate to identify or create a host for the data warehouse that fits the considerations described above.

The selection of an appropriate system administrator can determine the long-term health and direction of any neighborhood indicator system. The decision about system location should be made as early as possible, as part of an open process to build support for the system in its earliest and most vulnerable stages.

Data Selection

Systems in other cities provide examples of several approaches to deciding what types of data should be included in a data warehouse. These approaches, as described below, address general implementation and Detroit-specific considerations.

1. Use currently available data

Pro:

- This is the easiest and most cost-effective way to start and is a good first step. The completeness of the picture that the data present depends on the amount, variety and quality of data available. Currently available data for the Detroit neighborhood indicator system include all governmental data available to the public (census data and Home Mortgage Disclosure Act data), Detroit Finance Department - Assessments Division data and community-provided data.

Con:

- This approach excludes some relevant data that are might be considered crucial to a community-based organization, as well as data that are generated by neighborhood members themselves.

Currently available data for the Detroit neighborhood system include all governmental data available to the public (census data and Home Mortgage Disclosure Act data), Detroit Finance Department - Assessments Division Data and community-provided data.

2. Research and use the data described by literature, research, and current best practices as most effective for evaluating neighborhood health

Pro:

- Data used by researchers and practitioners have a greater chance of having been tested and analyzed for flaws, so this pool of data could generate strong indicators.
- This method works well for data used elsewhere for similar purposes, such as data needed for federal grant applications. For Detroit, voter registration, Section 8 Housing Choice vouchers, and Fire Department data are examples of data other systems use that are not currently available to P&DD or SDBA.

Con:

- Data used in other ways or in other places might not work well in Detroit neighborhoods because of different city characteristics.
- Data generated as part of a federally-funded study (for example, the National Institutes of Health grants) done at a major research university may not be readily accessible to general users because of protections for people who participate in health studies. (For a complete list of potential data to pursue see Appendix 1, Part III)

3. Gather data requested by users (institutions, individuals, community-based organizations)

A neighborhood indicators system in Minneapolis¹⁹ uses this method to gather data. The users of the system define their data needs, and the city and university partners try to meet those needs. This approach might also allow government agencies, foundations or other users to request that data be added to the warehouse.

Pro:

- Recognizes local knowledge by allowing residents to request data that meet specific needs in a neighborhood. This should allow for a more tailored fit between data and a specific neighborhood.
- This approach also ensures that the data warehouse includes data users most often need.

Con:

- Creates a more complex situation in which data are not comparable citywide.

¹⁹ Crossroads Resource Center. *Neighborhood Sustainability Indicators System*. <http://www.crcworks.org/nsip.html>. Accessed April 13, 2004.

4. Community organization-designed public input (citizens decide what indicators best represent their neighborhood).

This approach could also include neighborhood organizations collecting their own data.

Pro:

- The process is democratic.
- A more descriptive portrait of the neighborhood is usually obtained.
- This method is particularly useful to community organizations for program design and allocation of resources.

Con:

- This is the most time- and resource-intensive method of data warehousing.
- Data gathered this way require more staff time to ensure quality and consistency.
- The design and facilitation of surveys and data collection require extra funding.
- This method requires active community organizations in the areas for which information is gathered.

Conclusion:

All four approaches have merit for a variety of users: community-based organizations, foundations, city administrators, and the general public. All four could be included in a comprehensive data warehousing system. These approaches might be implemented in phases as follows:

1. Start a system using currently available data.
2. Conduct a study to identify relevant descriptive data for neighborhood health and find out which of them are currently unavailable.
3. Provide for data source research for acquisition of missing data.
4. Implement a system for public input into the warehouse.

Major Data Sources²⁰

After an intermediary decides which approach to take towards data collection, the organization will decide specifically what types of data to include in a warehouse. Listed below are several types of data that could be included in a data warehouse. The data are evaluated based on general implementation and Detroit-specific considerations.

²⁰ Several of these source categories were taken from Kingsley, G. Thomas, (ed). 1999. *Building and Operating Neighborhood Indicator Systems: A Guidebook*. National Neighborhood Indicators Partnership – The Urban Institute.

1. Census

Pro:

- Data from the U.S. Bureau of the Census (especially decennial census data) are inexpensive, easily comparable across space (nationwide or between census tracts), and cover a wide variety of information that is useful for looking at neighborhoods.
- The data include rich information about people and housing.

Con:

- Census data on population and housing are collected only every 10 years, so the information becomes dated quickly (even before it is released) and is not as useful for those interested in shorter timelines. This may become less of a problem in the future, when the Bureau of the Census provides data from the American Community Survey (see Conclusion below).
- Sample Census data are not available for geographies smaller than a block group.

Conclusion:

Census data are an important foundation of information in any data warehouse. However, a data warehouse should also include data that are updated more frequently. The American Community Survey may soon help to alleviate the problem of timeliness for census data, but only down to census tract-level data. According to the Census Bureau website, "...the Census Bureau will have annual data based on the American Community Survey once that survey becomes national in scope (July 2004). It will take three to five years to accumulate sufficient sample to produce data for areas as small as census tracts. Thereafter, tract data will be updated annually".²¹

2. Administrative records regularly updated and maintained by local agencies

Pro:

- These records contain a wealth of data about neighborhoods, including information about individual parcels or people. Voter registration, auto registration, library usage and school attendance may all be relevant to neighborhood health.
- Because they are collected at the smallest geography, these databases are often flexible. This means one can aggregate them to analyze small areas.

²¹ U.S. Census Bureau. *Question and Answer Center*. <http://ask.census.gov/cgi-bin/askcensus.cfg/>. Accessed April 5, 2004.

- The administrative data already are collected for a public purpose, so applying them to help people or neighborhoods remains true to that public purpose.

Con:

- Some agencies refuse to share data.
- Data collected for another purpose might be collected in a manner that makes it ill-suited for neighborhood indicator use or analysis. For example, sharing data raises confidentiality concerns if the unit of observation is the individual or if individuals can be identified. A data intermediary must be aware of this problem, and should take care to protect the privacy of individuals.
- Data collected from individual agencies may be hard to compare because of inconsistent methods across different data collectors. Agencies might use different units of analysis based on their own priorities, or collect data at different frequencies.
- It is difficult to inspect the validity of collected data. The data that agencies collect cover a broad spectrum of information. Verifying all of the varied data would likely prove cost- and time- prohibitive.

Conclusion:

Pursue collection of relevant agency data, but make sure the costs of using, updating and maintaining the data do not outweigh the benefit to the users of indicators.

3. Data purchased from vendors

Vendors such as DataQuick Inc. and Dun & Bradstreet keep family home sale prices and employment information that can be useful in analyzing neighborhood economic trends.²²

Pro:

- Firms collect data in a consistent and clear format that in many cases are updated annually.

Con:

- Data can be expensive and present a long-term, continuing cost to the data warehouse.
- One must evaluate the quality of data from each vendor to make sure they are reliable and valid.

Conclusion:

Vendor-supplied data might be avoided because of the expense and difficulty in verifying the accuracy of such data, unless the benefits of the data outweigh those factors. The cost of the data is potentially the largest consideration. Data might cost thousands of dollars annually, depending on the individual request and the company's policy.

²² Vendor data idea is discussed in: Galster, George, Chris Hayes and Jennifer Johnson. 2004. *Identifying Robust, Parsimonious Neighborhood Indicators*. The Urban Institute.

4. Special surveys or inventories

Data also can be gathered from surveys or inventories. These tools can provide both objective data (i.e., how many vacant lots on a street?) and subjective data (i.e., do residents feel safe in their neighborhood?). Surveys and inventories can be created and driven by community groups or conducted under the direction of a larger organization, such as a city government or a corporation.

Pro:

- Surveys can provide rich data not only about what is happening but what people think is happening. Neighborhoods often change because of perceptions, and understanding those perceptions might inform policies that affect neighborhoods.
- Some surveys and inventories also can involve or increase citizen input, or give residents an enhanced view of their role in understanding and changing their neighborhood. Those capacity-building attributes come in addition to the added data.

Con:

- Surveys are time-consuming and expensive, both in their development and execution. They also must be crafted with care to provide reliable, valid data.
- Surveys crafted by different neighborhoods might not be comparable with ones developed by other neighborhoods.
- Partnerships necessary to do community surveys also require money and time.

Conclusion:

Include neighborhood surveys to the extent practical, considering staff and budget limitations. The survey can be led by one entity with the participation of neighborhood organizations.

Clearly, each of these types of data can produce benefits within a data warehouse. Due to resource and time limitations, an intermediary might face certain challenges in acquiring the data. A discussion of some of these challenges and widely-used responses follows.

Data Acquisition and Management

Once the desired sources are identified, getting the data presents another set of challenges. Although one might find Census or vendor data easy to obtain, getting shared information from other agencies may be more difficult. This is particularly true of disaggregated data that create confidentiality concerns.

The National Neighborhood Indicators Partnership found three major obstacles to data sharing among agencies.²³

1. Providing data is too much work. This is largely unfounded because technology makes the transfer of large amounts of data easy. The city of Detroit already captures some data using regular data transfers. For example, the Detroit Assessor's Division of Finance Department provides annual updates to P&DD at roughly the same time each year. Once the initial data formatting and schedule challenges are sorted out, this process takes minimal ongoing effort on either party's part.
2. Confidential client information might be released to the public. This is a serious concern, but indicator systems find ways to assure confidentiality. The Boston Foundation, according to Kingsley, signs an agreement promising to safeguard information about individual families and:
 - Forbids access to the original file to anyone except the computer operators at Northeastern University
 - Makes public only data that are aggregated at the block-group level or higher
 - Suppresses block-group data when they are based on fewer than five observations.

Confidentiality agreements should be negotiated with each data provider so the data providers know their data will not be misused and will not violate the privacy of those people whose information is in the data. The degree of confidentiality should be discussed with the data providers. If necessary, data aggregation might be a good approach to avoid such sensitive issues.

3. They might be embarrassed by the data because the data could reflect badly on their performance. NNIP cities argued that the benefits outweigh the risks of embarrassment. The data warehouse can save agencies time and effort because they do not have to work to collect data from other agencies or have to deal with so many data requests from users. The data warehouse can enable many people to obtain the data they would otherwise request from an agency.

The most important role of city departments and agencies in the development of a neighborhood indicator system is to help provide the necessary tools for establishing the system.

Role of City Departments in an NIS

The role of city departments and agencies within a neighborhood indicator system is typically to help provide the resources necessary to run a successful system. The scope of their role

²³Kingsley, G. Thomas, (ed). 1999. *Building and Operating Neighborhood Indicator Systems: A Guidebook*. National Neighborhood Indicator s Partnership – The Urban Institute.

varies depending on the amount of interest the city takes in the system, the perceived benefits and the strength of partnerships and trust between the city and the system's administrator. The two main roles city departments serve are:

1. Data Source

A continued supply of data from the city is vital for a successful system to establish itself, grow, and provide a reliable data source for system users. For example, in Philadelphia the city provides data through the Mayor's Office of Information Services, Office of Housing and Community Development, Department of Licenses & Inspections, Philadelphia Gas Works, Revenue, Water Revenue, Philadelphia City Planning Commission, Philadelphia Housing Association, and Philadelphia Fire Marshall's Office.²⁴ Similarly, in Milwaukee the city provides data through the Milwaukee County welfare records and city property files.²⁵

2. Technological Funding and Support

Cities also provide technical assistance, staffing, and seed funding in the early stages of system development. While the seed funding is vital in the initial stages, providing data is the city's ongoing commitment to the system. Without initial funding the system will not be able to establish itself to raise more money or facilitate any research that would potentially lead to more data sources and more investment from the community. Depending on the resources available through the system administrator, the city may provide technical resources such as computers or training on how to use various computer programs.

The most important role of city departments and agencies in the development of a neighborhood indicator system is to help provide the necessary tools for establishing the system. As the system grows, the role of the city continues as one of the major data providers, though other parties eventually share the burden of data provision, funding, and technological assistance. The system administrator becomes more self-sufficient as more users and data sources are involved, but city-provided data will always be necessary.

Metadata and Data Management

Any data warehouse should have a strong foundation of metadata for users to understand what the information they have means. This section defines and explains why metadata is important.²⁶

²⁴ Cartographic Modeling Lab. *Project Profile*. http://cml.upenn.edu/project_areas/nis.htm. Accessed April 17, 2004.

²⁵ Employment and Training Institute, University of Wisconsin-Milwaukee. *Neighborhood Indicators for Central City Milwaukee 1993 – Present: Background* <http://www.uwm.edu/Dept/ETI/reports/indypage.htm>. Accessed April 17, 2004.

²⁶ For further reference, an in depth discussion about metadata is available from: ESRI. 2002. *Metadata and GIS*. <http://www.esri.com/library/whitepapers/pdfs/metadata-and-gis.pdf>. Access February 18, 2004.

Metadata is information about data that has been collected and compiled. It tells the user about what the different fields in a spreadsheet or a database mean, who compiled the data, when the data were gathered or for what time period the data apply. The presence of metadata allows the user to work with the data because definitions and terms are clearly explained. Without metadata users are often frustrated because they do not know what different codes mean or how different numbers have been calculated. The provision of metadata increases the efficiency and effectiveness for those who analyze data and for those who interpret the results from a study involving these data.

The creation and distribution of metadata streamlines the process of data gathering and sharing. The list below identifies some of the benefits of having comprehensive and current metadata associated with all information in a data warehouse.

1. Allow for easy data searching to expand or revise the scope of analysis

Users can query metadata to find data previously unnoticed but potentially relevant to their research question.

2. Help users easily evaluate data's relevance to their question

For example, if a study relies on data from a particular year, the project manager must check that the year of the data in the warehouse matches the year that is being studied.

3. Facilitate the selection of appropriate data from the warehouse

A model developer must know the definition of various data fields and how they were generated in order to use, calculate, and interpret the variables in correct ways. For example, the U.S. Census has a definition for "households" that differs from the definition for "families." Therefore, household information is different from family information.

4. Clarify which data need to be updated and when

Associating dates with all data ensures one is using the most current data. Knowledge of the consistency of data collection is also important, as consistency suggests that the researcher can rely on the data being used.

5. Improve understanding and ensure the study is replicable

Any findings from a study or model are useful only if they are reproducible. For a study to be conducted by an independent third party, its components must be readily understandable. Good metadata makes both of these possible.

6. Make studies and models more understandable for audience members

Metadata not only helps the user in constructing a study, but also the person trying to understand the study results.

7. Make more data accessible to more people

Metadata explains data, making the information understandable and useful, and therefore accessible to users with varying levels of familiarity with the data and the data warehouse itself.

8. Save time and money for both the warehouse administrator and users

Metadata helps manage data collection and documentation. When the importance or relevance of data is clear, one reduces the amount of money spent acquiring unnecessary data and the amount of time trying to find the desired data. Metadata also makes the process of working with data more efficient by easing the transition when a project is handed from one employee or organization to another.²⁷

Metadata Recommendation

Metadata is a key component of management for any data warehouse. This plan recommends that all those involved with this NIS, data acquisition, or projects involving data manipulation make recording of metadata a standard practice. If the city or any CBO has concerns about privacy, they should establish different levels of access to the metadata through a registration process or by establishing user agreements. If metadata cannot be obtained, the following is a list of the most basic metadata needed for data to be useful as a guide for staff of the data warehouse in development of metadata.

- Definition (what are the data)
- Derivation (how were the data calculated)
- Date (to what year, month or day do the data refer (this is not the date of collection))

For example, the indicator Serious Crime per Capita stands for the number of total serious crimes per 100,000 people, and is defined as the combination of violent and property crime. This is a measure that can be compared across census tracts because it accounts for differences in population. Serious Crime per Capita is derived by adding the values for violent and property crime (both per 100,000). The date for this indicator is 2002. The data refer to the number of crimes that occurred in 2002.

Conclusions about Data Acquisition and Management

Acquiring administrative data requires a combination of personal skill in working with those who create the data, political skill in convincing city leaders that the project is important and negotiation prowess to get data from reluctant agencies without

²⁷ ESRI. 2002. *Metadata and GIS*. <http://www.esri.com/library/whitepapers/pdfs/metadata-and-gis.pdf>. Accessed February 18, 2004.

Johnson, Steven. *The Quest for Data and the City of Orem, Utah*. <http://gis.esri.com/library/userconf/proc03/p0145.pdf>. Accessed April 10, 2004.

undermining their institutional needs.

Data acquisition is easier if political leaders convey that it is important. This can lead to an atmosphere of data sharing between agencies.²⁸ Working with the individuals who create the data can be fruitful. NKLA, the neighborhood indicators system used in Los Angeles and run by the University of California at Los Angeles, was able to collect data by asking agency workers if they wanted data from other departments or agencies. These data creators then wanted to share data in order to get data from other people.²⁹

Negotiation is important. In the case of the County of Los Angeles, which was reluctant to share data because it sold its electronic assessor database to private firms, NKLA reached an agreement to buy a portion of assessor data so as not to limit the marketability of the information.³⁰

Success encourages more data sharing. Agencies presumably will want to be part of a successful program and will see the benefits of acquiring data from a data warehouse.

Conclusion

A data warehouse is the foundation of the proposed Detroit Neighborhood Indicator System. To ensure long-term success, the system administrator should give time and careful consideration to issues of data management and control, data selection, and data acquisition. By making decisions about these areas in an open discussion among participating partners, the system can start from a strong foundation. In turn, the Early Detection and Community Empowerment System applications that draw on the data warehouse will more likely succeed in their goals of helping improve the lives of Detroiters.

To ensure long-term success, the system administrator should give time and careful consideration to issues of data management and control, data selection, and data acquisition.

²⁸ Kingsley, G. Thomas, (ed). 1999. *Building and Operating Neighborhood Indicator Systems: A Guidebook*. National Neighborhood Indicator s Partnership – The Urban Institute.

^{29, 30} Neighborhood Knowledge Los Angeles. *How-To-Kit*. <http://nkla.sppsr.ucla.edu/aster.cfm?Page=HowToKit/main.cfm&Page2=Political.cfm#2>. Accessed April 19, 2004