THE GEOARCHIVE HANDBOOK
A Guide for Developing a Geographic Database as an Information Foundation for Community Policing

ILLINOIS
CRIMINAL JUSTICE
INFORMATION AUTHORITY

THE GEOARCHIVE HANDBOOK

A Guide for Developing a Geographic Database as an Information Foundation for Community Policing

by
Carolyn Rebecca Block
and
Lynn A. Green

Statistical Analysis Center
Illinois Criminal Justice Information Authority
120 South Riverside Plaza
Chicago, Illinois 60606

The research described in this report was funded in part
by the Department of Justice, Office of Justice Programs,
Bureau of Justice Statistics,
Analytic Projects for State-Level Criminal Justice
Statistical Analysis Centers (SAC-2 program)

120 South Riverside Plaza
Chicago, Illinois 60606-3997
312-793-8550

Originally printed by the authority of the State of Illinois
August, 1994
Second printing February, 1996
Printing order number 96-067
250 copies

Contents

Maps

Acknowledgements

This report is based on the experience of the Early Warning System for Street Gang Violence project team, which included both Illinois Criminal Justice Information Authority staff and Chicago Police Department (CPD) Area Four Detective Division staff, and was supported by an Advisory Board of community experts and research specialists. (For a list of Advisory Board members, see Appendix I). The project was funded in part by a grant from the Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

The Chicago Police Department, particularly the Detective Division under Chief John T. Stibich, deserves much of the credit for the accomplishments of this project and thus, for the contents of this report. The project was initiated at the suggestion of the former Area Four Violent Crimes Commander, James A. Maurer (now a Deputy Chief in the CPD Detective Division), at a meeting with Project Director Carolyn R. Block and Loyola professor Richard Block (who serves as a member of the project Advisory Board). Deputy Chief Maurer has been a staunch, innovative and resourceful supporter throughout the project. Area Four Case Management Sergeant Ronald Rewers, former CPD Research and Development Division Commander Joseph Beazley (now Joliet Chief of Police), and former Gang Investigation Section Commander Robert W. Dart were also active in the initial development of the project and charter members of the Advisory Board. Sergeant Rewers continues to work closely with the Area Four Early Warning System, and the project is now benefiting from the support and encouragement of Violent Crimes Commander John Kozaritz, Area Four Detective Division.

Area Four detectives and civilian staff who were involved in the development of the Early Warning System were James Elliot, Tommy Waters, Lawrence Soltysiak and Richard Respondi. Detective Respondi, who used a computer for the first time for this project, has become an expert in spatial analysis and regularly conducts and extensive and detailed spatial analysis for the department. Detective Soltysiak is now serving as a technical advisor to the five prototype community policing (CAPS) districts, applying GeoArchive and spatial analysis technology to CAPS.

Advisory Board member Gregory T. Kowalec, Director of the CPD Data Systems Division, was instrumental in helping us confront data access and translation problems and developing the concept of the *GeoArchive Working Group,* and Data Systems Analyst Cindy Herzberger has been patient and persistent in solving data problems. The assistance with data problems provided by Lieutenant Phillip A. Chomiak of the CPD Research and Planning Division and Officers Hal Ardell and Lawrence J. Bobrowski has been invaluable. Officer Bobrowski, as the initial developer of computer mapping in the department, provided district boundary files to the project and Officer Jonathan Lewin provided boundary files of police beats. In addition, the project has benefited from the help of Advisory Board members Barbara McDonald, director of the Research and

Planning Division, as well as Chief Sherwood S. Williams, Lt. John Kennedy and Officer Art Block.

Because the essence of a GeoArchive is the combination of law enforcement data and community data, the role of the community members of the Advisory Board has been vital to this project. We gratefully acknowledge the contributions of Glenda Cunningham of Bethel New Life, Inc.; Robbie Jacquette, consultant; Joseph Guerrero, Supervisor of Dvorak Park; community member Sonia Gaete; Ann Cibulskis and Taya D. Sun of the Chicago Housing Authority; and Rey Colon and Christine Corrado of the Boys and Girls Clubs of Chicago. We also greatly appreciate the many local and state agencies who contributed much of the community data to the GeoArchive. In particular, we would like to thank John Karnuth, Senior Operations Analyst, Chicago Department of Housing, and Janice K. Suttie, Program Analyst, Division of Administration, Illinois State Police.

Three Advisory Board members are professors offering technical advice and direction to the project: Michael Maltz of the University of Illinois at Chicago, Richard Block of Loyola University Chicago, and Irving Spergel of the University of Chicago. Professor Spergel directs CPD's Street Gang Violence Reduction project in District 10 of Area Four. Michael Maltz contributed his expertise from an earlier Chicago mapping project, *Mapping Crime in its Community Setting* (Maltz, Gordon & Friedman, 1991). Richard Block, who pioneered automated crime mapping in Chicago (Block, 1977) and was the first to use the Chicago TIGER file map, has been instrumental in every stage of the project. He and his students provided much of the geocoding of the street gang data and created many of the boundary files that are integral to the GeoArchive (see the companion report to this Handbook, the *GeoArchive Codebook*). In addition, he volunteered many days of his time, sometimes on an emergency basis, to keep the mapping systems at the Authority and Area Four up and running.

Illinois Criminal Justice Information Authority Research Analysts who were part of the project team included Graham Taylor, Lynn A. Green, Robert Bennett Whitaker, Lynn M. Higgins, Anthony Mata, Michael Maly and Carolyn Block. Tony Mata, who lives in the Area Four neighborhood, played a central role in the first year of the project and continues to serve the project on the Advisory Board. Research Analysts Sheryl Marek, Peg Dabdoub and Antigone Christakos provided the support we needed to enable us to focus on this work, and the support of John Firman, former Director of Research and Analysis at the Authority, was critical.

The most vital contribution, however, was made by Lynn M. Higgins, an Authority Research Analyst who is now part of the Chicago Police Department's CAPS evaluation project. Lynn not only compiled the initial draft of the GeoArchive Handbook but her expertise enabled her to contribute essential information on computer systems, database management.

Introduction

This report outlines issues and problems inherent in managing disparate and numerous geographical databases for use in crime analysis and law enforcement decision-making. It also presents some strategies to resolve those problems. Specifically, this report is a guide to the issues that face the developer and manager of a GeoArchive.

A GeoArchive is a geographic information system (GIS, see *Glossary*) with the following characteristics:

- It contains address-based, neighborhood-level information
- It contains both community and law enforcement data
- It is easily accessible to local decision-makers on a timely basis

A GeoArchive database supports mapped analysis of neighborhood and crime data. It is a repository of community and law enforcement data, organized for use in crime analysis and problem solving. When combined with a community problem-oriented policing program, a GeoArchive can become an *information foundation for community policing*.

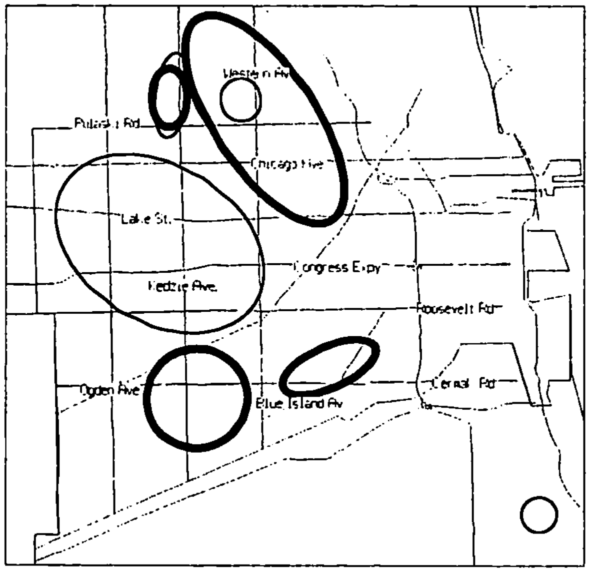
Crime maps, such as pin-maps on the wall inside a police station, are nearly as old as law enforcement. The new "high tech" version of these pin-maps -- computer-aided crime mapping -- gives police departments much greater capability to view crime data. However, the ability to view data is not enough by itself to support timely crime analysis and tactical decisions. To do that, we also need a well-organized body of local-level information (a GeoArchive) and the statistical tools to summarize that information.[1] Only when computer-aided mapping is combined with a GeoArchive and with spatial analytical methods can it become a powerful means to understand the crime problems faced by law enforcement and the community.

A GeoArchive is a *local* resource for crime analysis and decision making. In the past, computer mapping was so expensive and required such a high level of expertise that it had to be done centrally by a central administrative unit in the police department or even by a central city planning unit outside of the police department. This has now changed. The technological revolution in computer mapping means that maps and spatial analysis tools are available to analysts at the local, district and neighborhood levels.

This advance in crime mapping and data analysis comes at a propitious time -- when community policing is rapidly becoming the principal method of policing across the nation. A community's success in increasing safety and decreasing victimization often depends on its ability to pull together and interpret large amounts of information about a specific problem, and then to act on this information. However, the concept of community policing by itself does not provide this capability. For that matter, neither do high-tech tools such as computer mapping. It is the marriage of these two -- the community policing philosophy and the technology to understand crime patterns more fully -- that will yield more effective law enforcement in the future.

As an example of how analyzing mapped data can help law enforcement, Map 1 shows a spatial analysis of street gang-related crime in the study area of the Early Warning System for Street Gang Violence project (see Appendix II). Hot Spot Areas (the densest concentrations) of two gang-related criminal incidents are shown on the map: violent incidents (thick ellipses), and vice incidents (thin ellipses).[2] Knowing where specific kinds of street gang-related activity are concentrated -- that one neighborhood has a drug problem, another a violence problem, and a third has both -- can support the development of interventions that fit the specific problems of the neighborhood.

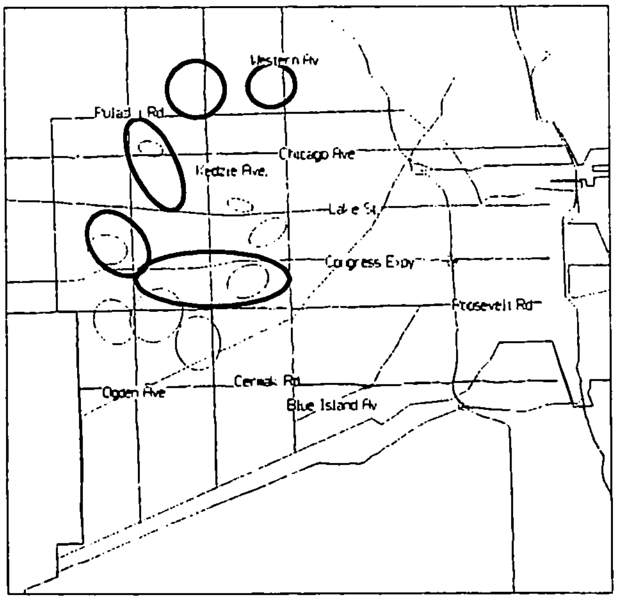
Map 1. Hot Spot Areas: Street Gang-Related Violent and Vice Incidents, EWS Project Study Area



Research has shown that successful interventions and strategies for prevention of property crime or violence require a two-pronged approach: first, identifying the problem, and second, targeting prevention efforts on that specific problem (for a review, see Block, 1991). To accomplish the first, we must organize and sift through the vast amount of information available about an area, each event anchored by location and time. Much information, such as street gang-related activity known to the police, the location of offenses and arrests, addresses of people arrested, addresses of people released from prison and citizen calls for service to the area, exists in law enforcement information systems. However, community information is also vital to a GeoArchive. For example, a pilot project utilizing spatially-related crime information in selected Chicago police districts (Maltz, et al., 1987) found that information from neighborhood and community groups added to the richness of the spatial database and allowed officers to identify high-activity areas more accurately.

Thus, a GeoArchive should contain not only law enforcement data but also community information. Creating such a database is the first step in developing an information foundation for community policing. Map 2 illustrates how law enforcement and community data can be used together for problem identification. The thick-lined ellipses show Hot Spot Areas of drug incidents in the Early Warning System project study area (Chicago Police Department Area Four); the thin-lined ellipses show Hot Spot Areas of abandoned buildings. Using this map, a crime analyst might decide that the apparent spatial relationship between some of the hot drug areas and the clusters of abandoned buildings should be investigated further.

Map 2. Hot Spot Areas: Street Gang-Related Drug Incidents and Abandoned Buildings, EWS Project Study Area



To get started in computer mapping, you must have two things. First, you must have a digitized map (see *Glossary*). This is a map with features such as streets, rivers or parks located in an x- and y- coordinate system. Second, you must have a way to place your data (criminal incidents, for example) on this map. This is called *geocoding* (see *Glossary*). A GeoArchive contains geocoded data together with the digitized map that forms the basis for the data.

There are numerous decisions to be made and challenges to overcome when developing a GeoArchive. Some of the most basic decisions will be the choice of data to be included and the best sources for that data. Another major challenge is how to link these data in a meaningful way and how to establish procedures for their use, while contending with the issues of accuracy, data sharing, data access, disaster recovery and privacy. This report is a general guide to the issues inherent in developing a GeoArchive.

A companion report, *The Early Warning System GeoArchive Codebook* by Lynn Green and Robert Whitaker (1994), provides more information about the GeoArchive developed for the Early Warning System for Street Gang Violence project, including details of the data source and GeoArchive histories of the databases and code specifications and definitions for all of the law enforcement, community and map data files currently residing in the GeoArchive.

This report is divided into the following sections:

- The first section, *Data for Crime Analysis*, discusses sources for law enforcement and community data, the kinds of information within databases, and the relationships between these data.
- The second section, *Data Verification*, outlines the principles of data verification. Data validity and integrity are of central concern in crime analysis.
- The third section, *Data Management*, covers the challenges of managing multiple, related databases and recommends standards and procedures for developing and maintaining a GeoArchive.

The document includes Appendices, a list of Advisory Board members and a brief overview of the Early Warning System for Street Gang Violence project and ends with a Bibliography with references to source material, a Glossary of key words and phrases and an Index.

Notes

[1]For an overview of spatial statistical tools, see Block (1994).

[2]Hot Spot Areas are calculated with the Authority's STAC (Spatial and Temporal Analysis of Crime) software (see *Glossary*). A Hot Spot Area is the densest cluster of events on the map and is defined by a standard deviational ellipse (see *Glossary*).

Data for Crime Analysis

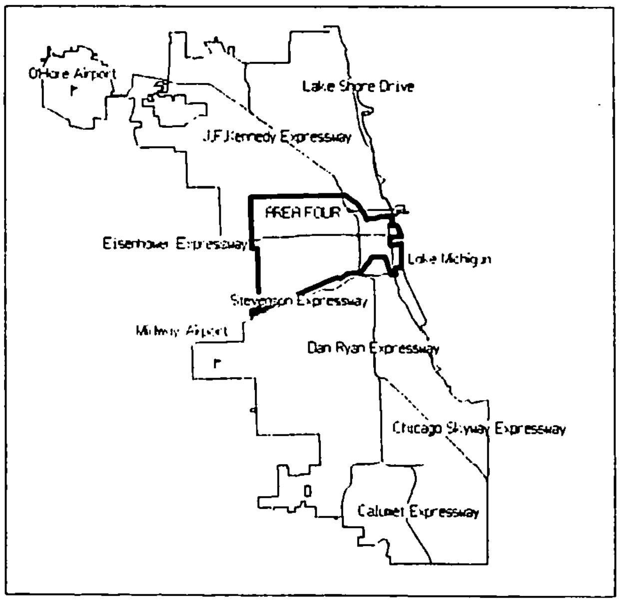
This section provides a general description of types of databases and data sources that might be included in a GeoArchive, organized in two general categories: law enforcement data and community data. For details about each dataset contained in the Area Four GeoArchive, see the *Early Warning System GeoArchive Codebook* (Green & Whitaker, 1994).

A GeoArchive contains geographical data together with the map that forms the basis for that data. There are three primary kinds of geographical data: points, areas and lines. The assignment of a set of coordinates to an address (point) or to an area is done through a process called geocoding (see *Glossary*). The base map for geocoding and mapping this information may contain such features as streets, rivers, expressways, etc. In this section, we describe the basic street map and geographic data files that make up a GeoArchive.

Point datasets include one or more data points. For example, in a point dataset of crime incidents, each point represents a single location on a map and is specified as x- and y- coordinates, latitude and longitude or through some other method of spatially placing a point. These points may be represented as different symbols or icons (see *Glossary*). In Map 3, the O'Hare and Midway airports are shown as mapped points.

In area data, a single data unit represents an expanse (however large or small) of geographical space, such as Census areas, parks, or street gang territories. A boundary (see *Glossary*) is a line enclosing such an area. Map 3 shows the boundaries of two areas, the City of Chicago and, within it, Chicago Police Area Four. Lines also are used to represent streets, railroad tracks, highways and rivers. For example, Map 3 shows the major Chicago expressways, represented as lines.

Map 3. Points, Areas, Lines: Chicago Airports; Chicago and Police Area Four Boundaries; Major Expressways



Points can be aggregated to areas. For example, the total number of criminal incidents in a Census tract can be easily counted. Areas, however, cannot be disaggregated to points. If we know only the total number of incidents in each police district, we cannot determine the location of each incident. Similarly, population information is area-level.

In the GeoArchive, both point data and area data have information behind them. They are more than dots or boundary lines on a map. For example, each crime incident in the Early Warning System for Street Gang Violence project has about 50 variables associated with it, including location, number of offenders and offense code; each Census tract has numerous demographic variables associated with it, including income and race.

Law Enforcement Data

The kind of law enforcement data available for use in a GeoArchive will depend upon the kind of data collected by or otherwise available to each individual department. Below we outline what law enforcement data could be included in a GeoArchive. We briefly describe each type of data, suggest possible uses, and list variables (fields; see Glossary) that might be associated with the data. Where possible, we provide an example or map from the Early Warning System project. For more detail, see the companion volume, *GeoArchive Codebook* (Green & Whitaker, 1994).

Incidents:

Data type: point

Incident data can be used to describe patterns of crimes such as auto theft, street gang violence, drug markets and burglary.

Variables may include information from the case report such as offense code; date and time of occurrence; location information (including the address, whether this incident took place on public transportation, a housing authority development, and the entry and exit points for the incident); the beat on which the incident occurred; street gang-related variables; the parties involved; weapon variables; and, situational aspects of the incident, such as whether there was an alarm or a fire.

The Early Warning System for Street Gang Violence project focuses on street gang-related crimes in a high-risk area of the city. The data are extracted from a city-wide incident data file, maintained by the Chicago Police Department, and records are selected on the basis of codes used to indicate a street gang-related offense.[1] Incident data are linked to victim and offender data via a record identification number.[2]

Arrests:

Data type: point

Arrest data provide information about the arrestee or about the arrest itself. They can be used to examine patterns involving a single offender.

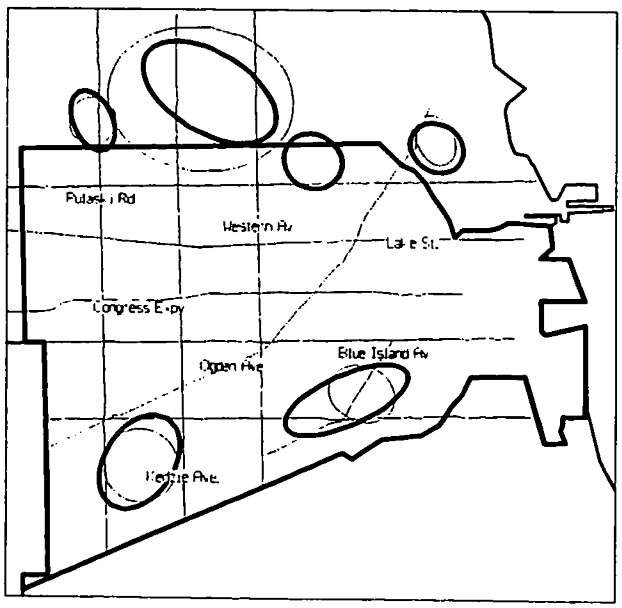
Variables may include demographic information about the person arrested (such as age, gender and race/ethnicity) and several different addresses (the person's residence, the arrest location and the offense location). Information about the incident, the victim(s), and associated incidents could be added to the file. In addition, variables may include how the arrest was made (investigation, at scene, etc.).

Offender data:

Data type: point

There are two types of offender data: data on suspects of specific incidents and investigation data or MO (*modus operandi*) data on known offenders. Incident-based offender data may include all suspects in the incident or only those found through police investigation to have committed the offense (whether or not they were arrested). When linked to incidents, offender data can be used to analyze travel to and from a crime (which is particularly useful in the analysis of street gang violence), drug markets, auto theft and auto recovery. Offender data also can be linked to *modus operandi* or investigation files, or to arrest data.

Map 4. Hot Spot Areas of Victims' and Offenders' Addresses: Violent Street Gang-Related Incidents, EWS Project Study Area



Variables may include information about the offender's prior record, gender, race, age, physical appearance and address.

In the Early Warning System for Street Gang Violence project, offender data are linked to incident and victim data to analyze the relationship between addresses of offenders and victims of violent street gang incidents (Map 4). The densest concentrations of victims' addresses (thick ellipses) do not always coincide with the densest concentrations of offenders' addresses (thin ellipses) in Map 4.

Victim data:

Data type: point

Victim data also can be of two types. First, incident-linked victim data can provide information about crime patterns relative to types of victims, the relationship of victim to offender (if any) and other incident-related analysis problems. Second, victim survey information can identify problems in a community for use in community policing problem analysis and tactical decisions. (Also see *Cognitive maps*, page 19.)

Variables may include information about the victim's gender, race, age, address, occupation and whether they were injured in the incident.

In the Early Warning System for Street Gang Violence project, victim data are linked to incident and offender data to analyze the relationship between the offenders and the victims of violent street-gang incidents or to analyze the addresses of the victims relative to the location of the incident.

Calls for service:

Data type: point

Calls for service data may be useful for understanding the level of activity in a community, since not all calls for service result in an official incident. In fact, many calls for service are not crimes at all but could be ambulance calls, fire department calls or calls to rescue a cat in a tree, for example. Calls for service data that may be useful include citizen reports of fights and loitering. Depending upon a particular department's recording of calls for service data, they may be linked to incident data.

Variables may include caller information, reported activity, unit assigned and disposition.

Probation release:

Data type: point

Data about offenders released on probation or on bond (pre-trial release) can provide information about the presence in a community of convicted or accused offenders. This information can assist in crime pattern analysis.

Variables may include probationer's name, address, sentence or bond type and conviction or charged offense.

Corrections release:

Data type: point

Data from corrections centers about released persons, including those on parole and supervised release, can assist in analysis that focuses on pattern development in a way comparable to that of probation data.

Variables may include releasee's name, address, time served, committing offense and parole or release status.

Street gang territories:

Data type: boundary (area)

Information about the turf of street gangs, especially disputed turf boundaries, is useful in analyzing activity relative to type of crime, type of offense and the victims and offenders involved in the crime. There are limits and issues with the use of territory files that are discussed below (see *Data Management: Street Gang Territories*, page 30).

Variables may include the name of the street gang, the nation (People versus Folk) and the street gang's offense history.

The CPD (Chicago Police Department) Gang Investigation Section provided information about street gang territories to the Early Warning System for Street Gang Violence project. Reliable sources in the community or on the police department are necessary to obtain this highly sensitive information. Further, the information about street gang territories changes rapidly and requires frequent revision to be useful.

Recovery of property:

Data type: point

Data about the locations where property is recovered can be used to locate chop shops, fences and others engaged in commerce in stolen property.

Variables may include address and type of property recovered.

In the Chicago Motor Vehicle Theft Prevention project, the project goals and targets were developed, in part, by an analysis of the locations where vehicles were recovered. This information is used to help locate chop shops.[3]

Nuisance addresses:

Data type: point

Information about locations that are considered by residents, business owners, police or others to constitute a nuisance or a danger to a community can be used to develop law enforcement and community strategies for abatement. These might consist of crack houses, street gang hang-outs or multiple-problem buildings.

Variables may include property information, complaints and previous arrests.

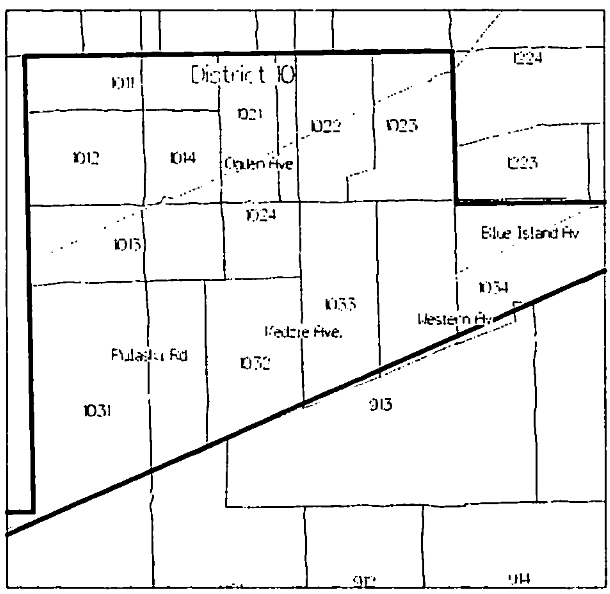
Criminal justice jurisdiction areas:

Data type: Boundary (area)

Most governmental agencies compile information within geographic boundaries. In Chicago, for example, the police department has areas, districts and beats. Other agencies, such as the courts, have their own districts.

It is important to include this information in a GeoArchive, for two reasons. First, many agency or community decisions may be specific to a particular kind of area, such as a beat. The GeoArchive should be able to support analysis that answers the agency's specific questions. In addition, sometimes it is necessary to relate data from one agency to another, such as police data to court data. If the GeoArchive contains the geocoded addresses of incidents and the boundaries of police and court areas, police incidents may be aggregated to court areas, and vice versa.

Map 5. Police Beats, CPD District 10.



Variables may include the name of the area (District 10; Marquette) aggregate data on the area (total incidents; population) or names and phone numbers of key people in the area. Map 5 shows the police beats within Chicago Police District 10. In the Early Warning System for

Street Gang Violence project, analysts aggregated data within beat boundaries to identify experimental and control beats for a Violence Reduction: Street Gang intervention project.

Community Data

In this section, we discuss some of the community data that are useful in building a GeoArchive, including street map data, land use data, public facilities data, and community characteristics data. For more detailed information about data in the Early Warning System GeoArchive, see Green and Whitaker (1994).

Street map data:

Data type: line (map)

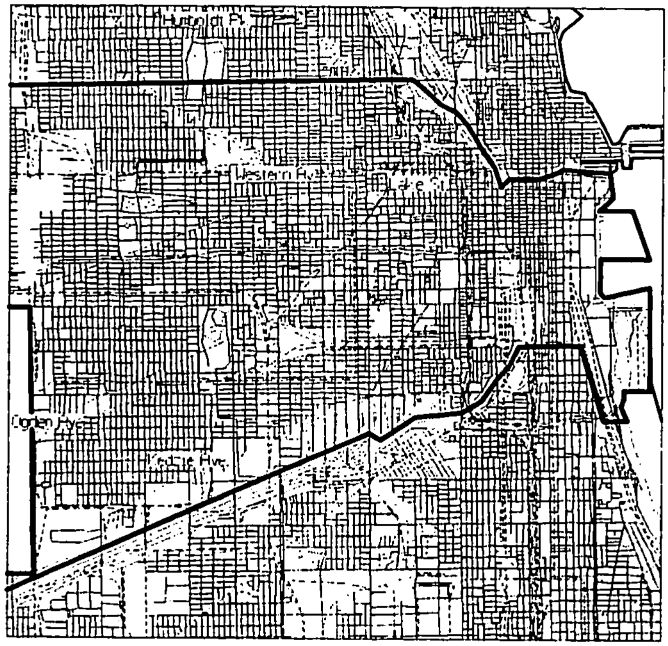
There are two things necessary for computer mapping: a computerized map and a way to place information on that map. To develop a computerized map, you must assign x- and y- coordinates to the streets, landmarks or other features on the map. This process is called digitizing (see *Glossary*). Then, to place information on the map, you must assign an x- and y- coordinate to each piece of information (such as the location of each criminal incident). This is called geocoding (see *Glossary*).

The accuracy of street map data is very important for two reasons. First, the map itself should be as accurate as possible. Second, the accuracy of the geocoded data depends upon the accuracy of the digitized street map data. There are computer programs that automatically geocode point data by matching the address of an event to the digitized street map. If the base map is in error, the program may not be able to locate an event, or even worse, it may put an event in the wrong location. Erroneous maps are, therefore, a serious threat to making accurate crime analysis or tactical decisions.

In the past, obtaining a digitized map was so expensive and difficult it was an obstacle to creation of a local-level GeoArchive. In the United States, however, the Census Bureau digitized the streets in every county in the country for the 1990 census. These TIGER files (see *Glossary*) are available from the Census or from computer software vendors. The availability of TIGER street maps is one of the main reasons that computer mapping is now a reality for small police departments and neighborhood-level districts. However, the TIGER files do contain some errors. There may be missing streets, either because new streets have been added to the area or because the original map was erroneous. In addition, there may be missing, misspelled or inaccurate streets.[4]

Because map accuracy is so important not only for mapping but for geocoding, it is vital to develop a system for correcting and updating street maps. Even street map data files for small cities and neighborhoods are large and complex (Map 6), and the maintenance of accurate maps is not simple. For a detailed discussion, see *Data Management: Street map data* (page 31).

Map 6. Streets: EWS Project Study Area



Property information:

Data type: point

In Chicago, the Sanborn Field Survey produced by the Chicago Property Information Project contains information about the status of property in the city.[5] In addition, a source of abandoned building information is Demolition Court Cases. (See Green and Whitaker, 1994, for details of both datasets.)

Examples of some of the many variables that may be available on such datasets include gross commercial floor space, total housing units, whether a parcel of land is used for parking or as residential quarters, abandoned building status, vacant lot locations, taverns, high-rises, commercial status and factories.

Information about abandoned buildings can be useful, since they often serve as sites where street gangs hang out and as drug houses.

Liquor license locations:

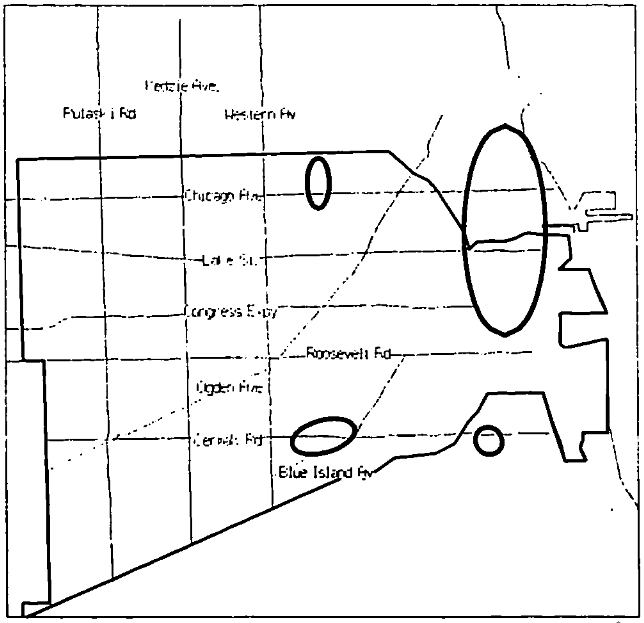
Data type: point

Data about liquor licenses can be used to identify troubled areas and nuisance places associated with the sale and use of alcohol. In addition to liquor stores, a bar, restaurant or tavern may be part of an environment that results in either violent or property crime (see Roncek & Bell; Roncek & Maier, 1991).

Variables may include the date the liquor license was issued to the establishment; the kind of establishment (tavern, beer garden, music and dance club, packaged goods store, restaurant that serves alcohol, etc.); the name and location of the establishment; and, the period for which the license is issued.[6] Incidents occurring at or near an establishment may be added to the file.

Analysis in the Warning System for Street Gang Violence project frequently found liquor licenses to be concentrated in the same areas as street gang "turf" battles. The highest concentrations (Hot Spot Areas) of liquor licenses in the EWS Project Study Area are shown in Map 7.

Map 7. Hot Spot Areas of Liquor License Addresses, EWS Project Study Area



Public transportation:

Data type: point, boundary (area) or line

Public transportation data, such as train and bus stops and routes, may be useful in identifying areas where high traffic is associated with high levels of criminal incidents. Sources of these data can be published materials of the transportation department.

Variables may include the name and destination of the route or stop. Details of associated criminal incidents may also be added to the public transportation file.

In the current GeoArchive of the Early Warning System for Street Gang Violence project, EL stops are mapped as point files. An alternative procedure would be to map each stop as a boundary file, which would provide a better description of large stations that span more than one intersection.

Schools:

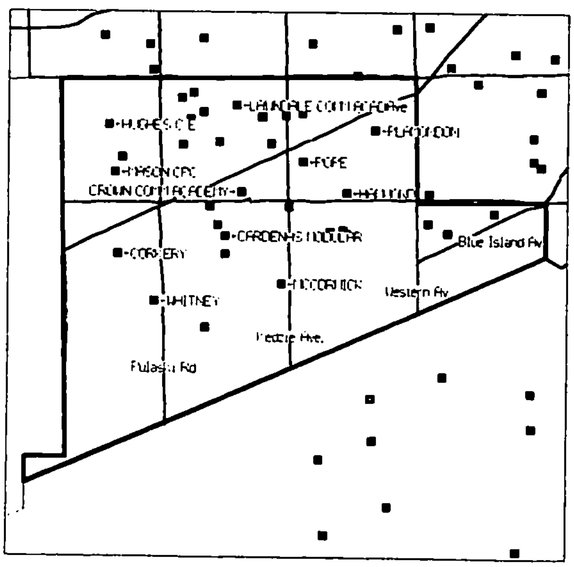
Data type: point or boundary (area)

Grammar and high schools are increasingly sites of a variety of crime both within the schools and in the surrounding area (see Roncek & Lobosco, 1983; Roncek & Faggiani, 1985).[7] Public school boards can provide information about public grammar and high schools. Information about private schools also can be useful. This information can be obtained through personal knowledge of a community or through organizations with which private schools are affiliated.

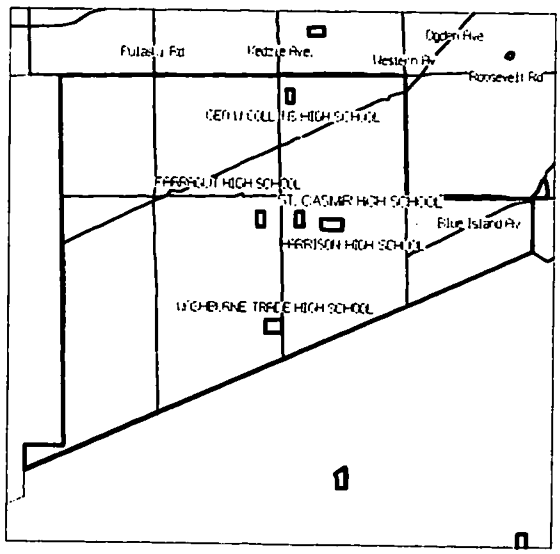
Variables may include the name and type of school, associated incidents, and important temporal information (such as the hours of dismissal or hours of Saturday-night games). Associated incidents may be those occurring within school grounds or within a given radius of school grounds and may include both police-recorded incidents and those reported by the schools or community members, such as vandalism. In addition to point data, boundary (area) maps may be created for larger schools, showing the school campus with each building and major entrances and exits.

In the Early Warning System GeoArchive, point files of public grammar school locations (Map 8) and boundary (area) files of all high school campuses (Map 9) have been added to the database.

Map 8. Locations of Public Grammar Schools, CPD District 10



Map 9. Locations of High Schools, CPD District 10



Community organizations:

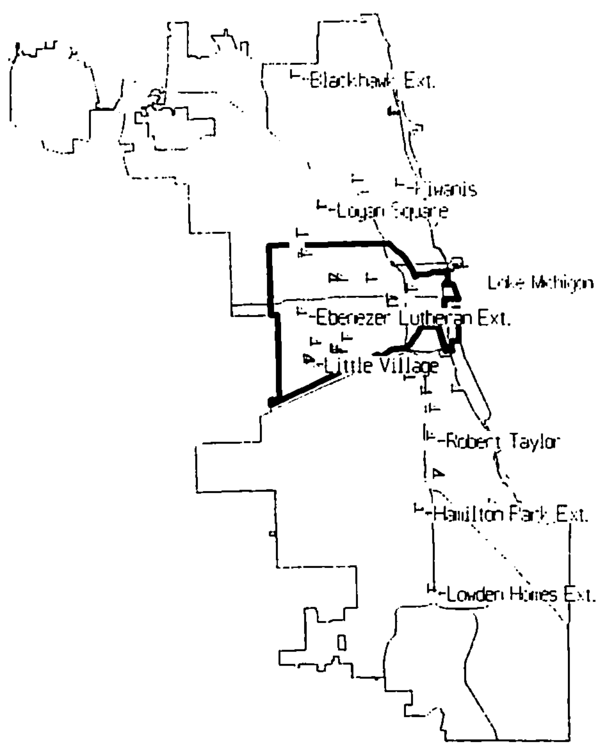
Data type: point

Community organizations of all kinds, such as block clubs, churches and social service agencies, are likely partners in community policing. They are often able to assist in problem identification and to contribute to problem solution.

Variables may include the name and type of organization, contact people with address and phone numbers and information about any associated incidents.

The Chicago Boys & Girls Clubs have been active on the project advisory board of the Early Warning System and in programs to reduce violence. Map 10 locates Boys & Girls Clubs across Chicago.

Map 10. Locations of Boys & Girls Clubs, Chicago



City parks:

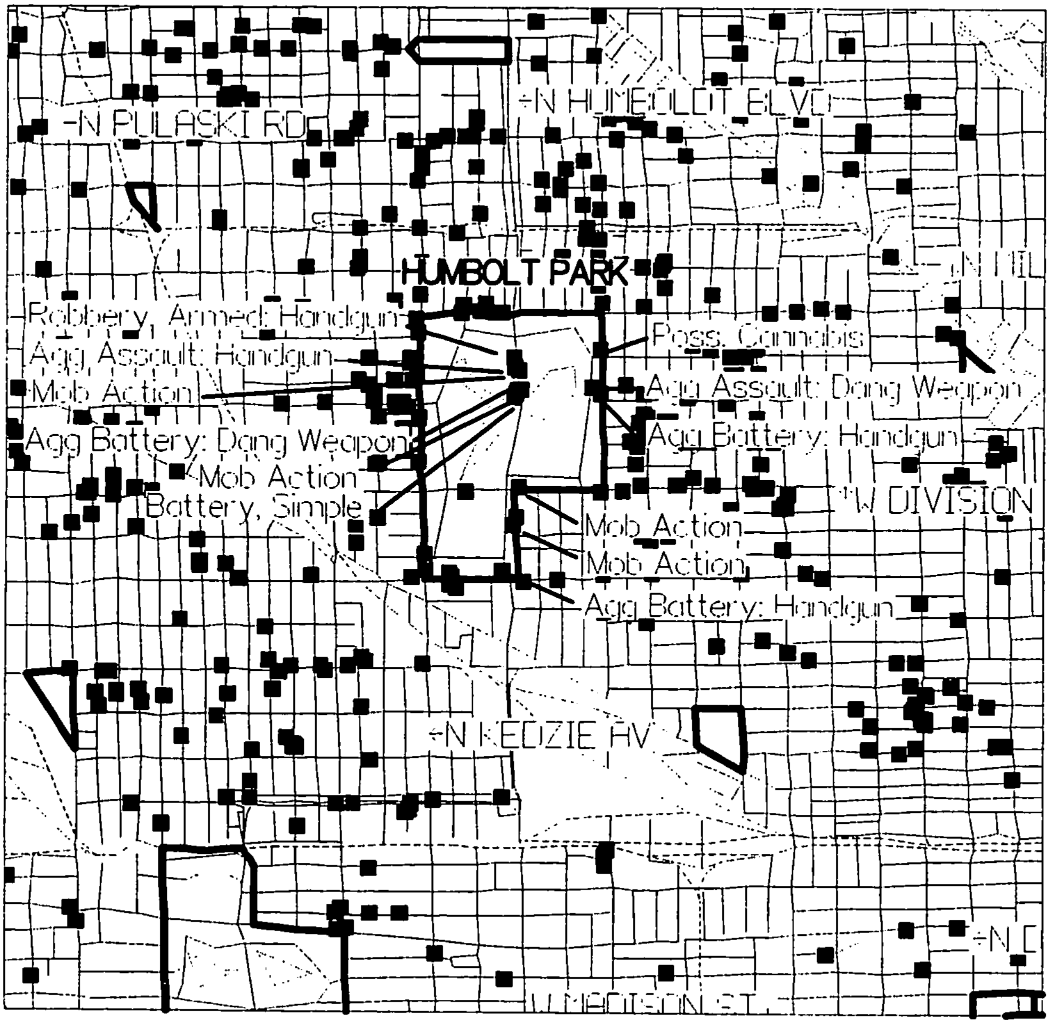
Data type: boundary (area)

The location of parks and other open areas can be an important factor in tactical investigation or problem analysis. Information about the location of city parks can be obtained from local maps.

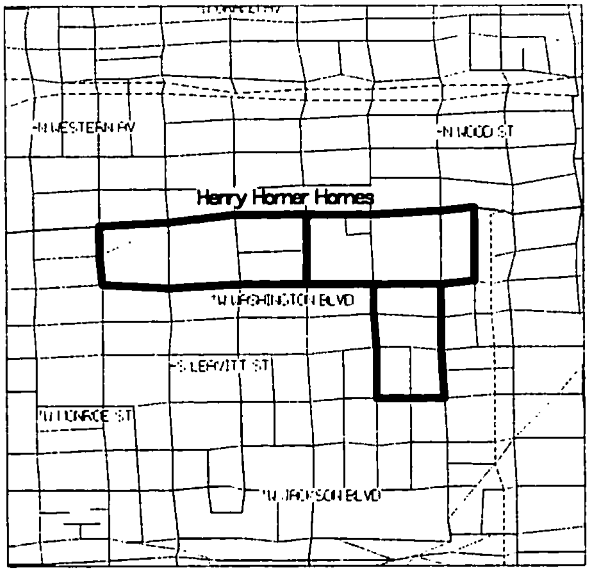
Variables may include the name of the park, the street gang territory and associated incidents (both police records and victim survey data).

Automated street maps like the TIGER files (see *Glossary*) only include address information for officially-defined addresses. An event that does not occur at such an address (for example, events within parks or parking lots, on a park street such as the driveway into the Planetarium, on park walkways, along a lagoon or the lakefront, in field houses, etc.) could not be geocoded with an unmodified TIGER file street map. To produce maps that include these vital locations, we create a modified version of the TIGER file (see Green & Whitaker, 1994 and Richard Block, 1994 for detailed instructions). With these modified street maps, we can place mapped incidents at the specific spot in the park where they occur and not at some arbitrary central point or intersection. Map 11 shows an example of this, street gang-motivated incidents that occurred in and around Humboldt Park, with some of the violent incidents labeled. Notice that incidents are specifically placed within the park at the location where each one occurred.

Map 11. Street Gang-Motivated Incidents, Humboldt Park Area



Map 12. Chicago Public Housing's Henry Horner Homes



Fire departments and police stations:

Data type: point

Fire and police department locations can be factors in crime analysis. For example, in those communities where community policing uses small neighborhood outposts, the effect of presence on activity can be analyzed. Locations of fire department stations usually can be obtained through the headquarters of the fire department or through the telephone book. Similarly, locations of area, district, beat and other police offices usually can be obtained through the headquarters of the police department or through the telephone book.

Variables may include the name and type of the facility. Associated fire department incidents or police dispatch information may also be added to the file. For example, fire alarms and calls could be compared to locations of arson incidents (in the police incident data) or to abandoned buildings.

In the Early Warning System project, the GeoArchive includes all fire stations and police district and area offices.

Public housing:

Data type: boundary (area) or point

Although a large body of research indicates some relationship between the presence of public housing and crime (see Roncek, Bell & Francik, 1981; Roncek, 1992; Dubrow & Garbarino, 1989), only limited data are available on the kinds of public housing situations that are particularly vulnerable. Collecting such information in a GeoArchive could provide valuable information for community problem-solving.

Variables may include the type and name of the housing unit, associated incidents and the specific kind of place (park, play lot, school, administrative facility, residence by type).

The Chicago Housing Authority (CHA) supplied information about the location of housing authority complexes to the Early Warning System project. In the current GeoArchive, each CHA site is one area. However, each of these areas may contain multiple city blocks in which there are multiple buildings (Map 12). Working with CHA staff, we hope to produce more detailed maps (see Green & Whitaker, 1994).

Additional considerations with mapping public housing and other kinds of complex structures have emerged in GeoArchive analysis, including maps of three-dimensional space (such as each floor of a high-rise building) and maps of the private roadways in CHA complexes. These roadways and other spaces require the same sort of street map modification discussed under city parks (*City parks,* page 15).

Population information:

Data type: boundary (area) or point

U. S. Census data are useful for calculating crime rates for specific kinds of crime within a block, block group, Census tract, police district or Hot Spot Area. (For details, see Green & Whitaker, 1994a, 1994b, and Richard Block, 1994.)

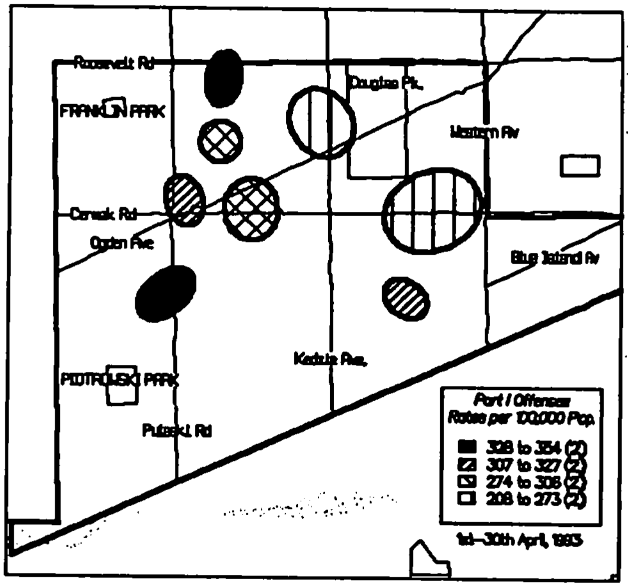
In addition, Census data are useful for analyzing the relationship between crime and characteristics of a neighborhood. This requires careful use of population and other aggregate-level information. For example, many models of "neighborhood disorganization" include changes in racial/ethnic residential patterns in their measure of disorganization. However, as Irving Spergel (1984:210) points out, "We also need to clarify the notion of community instability and distinguish it from racial/ethnic change." (George-Abeyie (1981: 108-109) and Roncek and Block (1983) also support Spergel's observation.)

Every record in Census data is related to an area such as a block or group of blocks (Green & Whitaker, 1994a). Area-level variables include information about population size, ethnic composition and level of income or poverty. In addition, every Census record contains the latitude and longitude of a central point (centroid) of the area (block or other area defined by the Census). This makes it easy to link Census information with other point or area data sets.

This sort of linking has been done in the Early Warning System project GeoArchive (see *Street map data*, under *Data Management*, page 31). For example, the density analysis shown in Map 13, the rate of Part I crimes per 100,000 residents, was calculated by dividing total incidents within each Hot Spot Area by the population residing in that area. The population within each Hot Spot Area was estimated by aggregating the populations of the Census blocks in which the centroid (see *Glossary*) of the block falls within the

Hot Spot Area. This analysis shows that the densest concentrations of Part I crimes in the study area are generally in the north or southwest.

Map 13. Part I Offenses: Hot Spot Areas and Rates per 100,000 People, CPD District 10



Public health data:

Data type: boundary (area) or point

Public health data can be useful in analyzing community characteristics. These data can be obtained from city, county or state public health departments.

Variables may include rates of mortality or morbidity, by specific cause, within a Census area (see above). The data may be linked to other datasets by using geographic information (the latitude/longitude of the centroid of the area). (See *Glossary* and *Population information*, page 18.) In addition, it may be possible to obtain specific addresses of cases of serious health problems, such as fatal firearm accidents and suicides, that may be very useful in an Early Warning System predictive model.

Mortality rates by cause of death, such as death by violence, can be used as a measure of community safety. Other community problems can be evaluated using indicators such as infant mortality rates and morbidity rates. Morbidity rates for diseases such as tuberculosis or AIDS can be used as indicators of community well-being.

Cognitive maps:

Data type: boundary (area)

Maps of community concerns and the resident's perceptions of community safety (cognitive maps) can be developed from surveys of neighborhood residents or of the community workers or police officers who work in a neighborhood. In addition, community groups may point out dangerous areas. Such information provides an additional perspective to official offense or arrest data. For example, at CAPS (community policing) beat meetings in Chicago, residents are frequently told they are the "eyes and the ears" of the police department because the officers cannot be everywhere.

at once.[8] Such information is not only useful to police investigation but also for crime analysis and problem-solving.

Variables in these cognitive maps of a community's social geography may include street gang territory locations, dangerous places (street corners or buildings), vandalism problems, drug market locations, etc. (Note that perceptions of street gang territories (see discussion below) are a type of cognitive map.) This information can be compared with locations of Hot Spot Areas of crimes known to the police and used in conjunction with both crime analysis and community decision-making.[9]

The individual cognitive maps obtained from a telephone or interview survey can be aggregated, showing the most dangerous regions or street corners in a neighborhood, as perceived by neighborhood residents. The effect of an intervention program (such as nuisance abatement or a community policing initiative) can be measured by before-and-after cognitive map surveys.

Notes

[1]See *Managing Data with the End User in Mind*, page 33, and the report, *Early Warning System GeoArchive Codebook* (Green & Whitaker, 1994), for more detail.

[2]The Chicago Police Department's Data Division provides regular "downloads" of incident data to the Area Four Early Warning System for Street Gang Violence project, from its mainframe database (see Green & Whitaker, 1994).

[3]Contact P.O. Jonathan Lewin, Research and Development Division, Chicago Police Department, at 312-747-6212 for additional information.

[4]See Richard Block (1994). for discussions of errors in the Chicago TIGER files.

[5]The Sanborn Field Survey is a continuing observation ("windshield survey") of every parcel of land in the City of Chicago, with a complete update to the file every two years. Inquiries can be directed to John Karnuth, Senior Operations Research Analyst, City of Chicago Department of Housing, 318 South Michigan Ave., Chicago, Illinois 60604. For more detail, see Green and Whitaker (1994a).

[6]Inquiries for Chicago data can be directed to: Janice Suttie, Programmer/ Analyst, Illinois State Police, Division of Administration, Information Services Bureau, Data Systems.

[7]See Illinois Criminal Justice Information Authority (1991) for information about crime and law enforcement in Illinois schools.

[8]Observations of beat meetings in a Chicago Alternative Policing Strategy (CAPS) prototype district by Lynn Higgins, October through November 1993, in conjunction with the Northwestern University Center for Urban Affairs and Policy Research evaluation of the implementation of CAPS.

[9]See also Rengert (1994) for an analysis of differences between cognitive and objective maps.

Data Verification

The purpose of data verification is to discover and correct errors in data. Error is a simple fact of data collection. The goal should not be to completely eliminate error but to establish a systematic process under which errors will be noticed, defined, documented and corrected. According to Francis Bacon (quoted in Kuhn, 1977), "Truth emerges more readily from error than from confusion."

There are numerous places where error can be introduced into a dataset. Sources of error include mislabeled files or databases, incompleteness and inconsistency and changes in the definition of fields and their values. Below, we discuss each in turn. Finally, we discuss how to handle erroneous data.

Mislabeled Files or Databases

Frequently, files may be moved from one department or agency to another. With some kinds of data, such as incident data over multiple years, it is also likely there will be multiple files or databases. Files may be delivered in a text format and loaded into database software locally. All these situations generate multiple, similar-looking files that may be confused with one another.

A priority of GeoArchive management is to assure that users always identify a file or database positively, even when they are in a hurry. This, we have found, can be a tremendous and serious problem in GeoArchive maintenance. If a GeoArchive receives regular refreshes (updates to the data, including updates and corrections to earlier years as well), the new files will supersede the old. Also, it is common to enhance a data file by geocoding (see *Glossary*) or by adding corrections or user-friendly summary recodes, for example. The enhanced data file will look essentially the same as the old file, but the difference between the two represents hours of work. An additional consideration in coordinating local changes with central GeoArchive datasets is that many people may make changes to the same dataset.

All these situations generate multiple but similar-looking datasets. In such an environment, it is easy for an analyst to load an outdated file, then by keeping it, write its data over the most current file, thus destroying a lot of work and creating confusion. To avoid this, we recommend the following rules for documenting data files:

- Document all data files, distinguishing between source and text files (files as they are received from the external department or agency source) and files that have been imported and manipulated by you.

- When documenting data files, it is a good idea to include the date of creation or updating in your documentation, to write the date on the disc label and to include the date in the name of the file.

- When documenting data files, distinguish the procedures for receiving and for implementing file updates.

- Document the relationship between the source files and any manipulated files. For example, document specific codes and recode rules for fields that have been added, deleted or combined.

- Maintain a codebook that includes details of the logical structure (file structure; see *Glossary*) for the database and the contents of each field (variable; see *Glossary*). (For an example, see Green & Whitaker, 1994.)

- Use standardized names for files and fields. To help distinguish between multiple versions of the same data, it is helpful to include the creation date as part of the name. For example, an earlier version of the Sanborn file might be labeled SAN91, and an updated version, SAN92.

- Back up your files and databases regularly. Backup to tape is efficient, and the purchase of a tape drive can be cost-effective. Store these backups off-site, if possible. While reconstruction of a destroyed file is possible, simple restoration procedures are usually far less difficult.

- Document how each database is used at your site. Maintain a flow chart showing the progress of each database and the naming conventions used at each stage. Distinctive names might be required, for example, as a data file is received from the source, is cleaned, geocoded and enhanced, is transformed into a user-friendly version, and, finally, is superseded by a refresh file from the source.

In general, these management practices reduce error resulting from mislabeled files and databases: using naming conventions and standards for loading and saving databases; storing original or source files in a separate directory or on floppy disk, preferably off-site; and, standardizing the import process so there is only one way to do it.

Incomplete or Inconsistent Data

For each database, it is necessary to determine whether the data are complete and consistent. Databases that do not have missing records or cases are complete. Databases in which fields are defined and used consistently over time and across different jurisdictions are consistent. Assessing completeness and consistency can be done by determining whether the data contained in a database are reasonable.

Some rules of thumb for detecting erroneous data include the following:

- Examine the number of records or cases in each database. Check to see that the total number of records or cases that you have is the same number as the data source reported sending to you.
- Ask whether the numbers (frequencies) look right for each code in key categories. In other words, do they pass the "giggle test?" Frequencies and totals can be compared with source documents, such as published statistical summaries.
- Look especially for codes that are not used, codes that were used in one time period and are not used in a later time period (new and discontinued codes), and missing data.
- Plot data over time -- monthly, weekly or daily, depending upon the appropriate level for the categories and codes with which you are working. If numbers in a time period are much higher or much lower than previous time periods, find out why.
- Look for outliers (data patterns that do not fit the norm for a category). An example might be an excessively large number of offenders in an incident. The case may be explainable, but it may also be an error. These should be investigated and marked or flagged as exceptional cases or corrected if an error has indeed occurred.
- Investigate any codes in use that are not in the existing codebook. Find out what the codes mean and correct the codebook. If the codes do not exist, mark them as mistaken coding or "missing data." If you have no information, flag the codes "undefined."
- If there are many undefined codes, suspect that the file structure (see *Glossary*) may be in error.
- Look for non-reporting (by district, time period, a field or type of information).

- If you receive data on a monthly (weekly or daily) basis, calculate the mean and standard deviation of your monthly (weekly, daily) totals as you receive the data. Given random data, you would expect that a month would be higher than one standard deviation above the mean only about 16 percent of the time, and the chance of being higher than two standard deviations is much less.[1] Investigate these situations.

Changes in Definition

Some data are part of a long history of data collection and maintenance. Literally decades of data may exist. Often, the meaning of variables may change as a result of changing needs or perceptions over time. These are not errors *per se*, but may easily result in misleading analysis or erroneous conclusions. Therefore, these definitional changes must be discovered and documented. Then, to avoid misinterpretation, the changes must be incorporated into the user's database.

If inconsistencies are noted in the process of verifying the data, check to see if the meaning of the field has changed. New codes come into existence; old codes go out of existence. For example, in 1990, CPD changed the definition of the code "prestige" to the more descriptive code "affiliation," capturing some cases previously included under other codes (see Green & Whitaker, 1994). The history of the data is a key to understanding data discrepancies over time.

Handling Erroneous Data

So far, this section has discussed how to identify errors in a dataset. Once an error has been found, it should be corrected if possible. If it is not possible to correct the problem, the situation should be documented for users of the dataset. In any case, steps should be taken so that the particular type of error does not happen again. In summary, the following actions can be taken to handle erroneous data:

- Be sure there was no communication problem in requesting or receiving the data from the original source. Examine the specifications for the request of the file. Was the "missing" field requested?

- Check for an erroneous file structure (see *Glossary*). Be sure you are looking in the right place for your "lost" fields or variables.

- Determine if there was a code or procedural change, either at the original data source or in GeoArchive management.

- Determine if there was a change in the practices of those units or divisions assigning codes to the field.

- Working closely with the original source, where appropriate, attempt to correct the error. Some simple errors, such as a misspelled street name, can be corrected locally, without consulting the original source. Others, such as changes from year to year in coding policy or practice, are more complex and require consulting the original source.

- Document all decisions made about changes to the file based on the discovery of error. In cases where the error was caused by GeoArchive management practices, poor communication with the source or an incorrect codebook, correct the problem to avoid future error.

To minimize error in data, the following principles should guide the standards and procedures you implement:

- Make sure verification procedures are easy to use by setting up "canned" programs where possible.

- Promote data collection by data users. Data will always be more accurate if the personnel responsible for data collection are also users of the data.

- Edit your documentation, codebooks and user manuals until they are clear and easy to use.

- Make your coding rules easy to follow and use.

- Provide continuous training that reflects job duties closely and is hands-on.

- Set up an easy and quick problem resolution process that is readily accessible to a coder or data management worker who has a question.

- Provide feedback about error discovery and correction to each user.

Note

[1]In conducting such an analysis, you should be aware that "time series" data are not independent. Therefore, such statistics as the mean and standard deviation should be considered *exploratory*. Also, be sure to allow for cyclical effects, such as seasonality, or for peaks at certain days of the week or on holidays. For more information, see Block (1984, 1987).

Data Management

The problem of deciding what data are best for a particular GeoArchive is not simple. Should you wish to maximize the amount of law enforcement and community data you have available on a day-to-day basis, you will need to establish relationships with numerous community and city agencies and departments. Once data have been obtained, you must establish and manage the GeoArchive. This is labor-intensive. In addition, you must plan for updating and protecting these data.

In this section, the problems of managing multiple databases and dealing with multiple data sources for the GeoArchive are outlined and discussed. Next, we discuss how to manage data for ease of access by the end user. We then recommend standards and procedures to assist in development and maintenance of a GeoArchive.

Managing Multiple Databases and Data Sources

The objectives of the development and maintenance of a GeoArchive are to have up-to-date, timely information while at the same time ensuring accuracy and validity of data. Sometimes, these two objectives cannot be met simultaneously. When this is the case, it is necessary to be cognizant of the trade-offs made and the resulting limits of the data.

Three kinds of problems in the management of multiple databases and data sources are outlined and discussed in this section. First, there is the problem of "linking" databases, that is, developing and updating databases that have relationships to one another and that can be used together for problem-solving and analysis. Second is the problem of rapidly changing and accumulating information. Third, there is the problem of synchronization between data source and GeoArchive.

These problems are discussed below in the context of specific examples:

- *Linking victim and offender data with incident data* illustrates the problem of linking databases.

- *Mapped street gang territory data* illustrate rapidly changing, complex and accumulating information. Street gang territories are completely local, differ according to perspective and change rapidly.

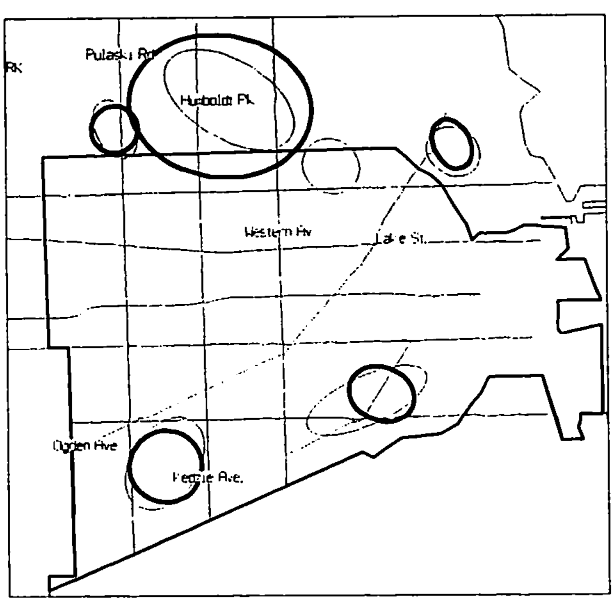
- *Street map data* are highly detailed, complex and prone to error. They illustrate the problem of synchronization between data source and GeoArchive.

Linking victim and offender data with incident data:

The Chicago Police Department (CPD) keeps files of information about criminal incidents and the victims and offenders in those crimes. To prepare incident, victim and offender files of street gang-motivated offenses for the Early Warning System for Street Gang Violence project, the CPD Data Division first selected street gang-related incidents (see Bobrowski, 1988, for details of the Chicago definition of street gang-related incidents). Then, the incident record number on the incident file was matched to the incident record number on the victim and offender files. Linked files were created by the Early Warning System analysts by merging the three CPD-created files.

With linked incident, victim and offender data, analysts can search for concentrations of victims and of offenders, search for possible patterns or relationships between them or analyze their association to the incidents themselves. Map 14 shows Hot Spot Areas of victims' and offenders' addresses of street gang-related incidents in and around CPD Area Four. The thick ellipses show the densest concentrations of offenders and the thinner ellipses are the densest concentration of victims. Such linked data allows analysts to explore issues like these:

Map 14. Hot Spot Areas of Victims' and Offenders' Addresses, EWS Project Study Area



- Are there neighborhoods where there are many drug offenders but few drug incidents?
- Are members of a particular street gang more likely to be offenders in an incident than victims?
- Are there neighborhoods where there are many street gang-related violent incidents, but where few victims or offenders reside?

As this example shows, linking incident, victim and offender data can enable law enforcement and the community to focus intervention strategies on the specific nature of the neighborhood's problem. Linked data can also help find hidden or unexpected problems in a neighborhood (as with the drug offender example, above). Coordinating the process and logic of data linking is no simple undertaking, however. File management is a serious issue since it involves creation of multiple files with different geocoded addresses.

A fundamental necessity of data file linking is that each file to be linked has an identification number that is the same across files. In the Early Warning System project, for example, each record, whether incident, victim or offender, has an "RD" (Records Division) number denoting the case. Although there may be more than one offender or more than one victim per case, all the offenders and victims can be matched to the proper incident by the RD number. Variables on the incident file that are not on the offender or victim file, such as the type of offense, can be added to the offender or victim file, because the RD numbers can be matched. If you do not have some ID number that exists in all of the data files you want to link, it may be impossible or very difficult to link your files.

One way you can link files without a matching ID is through computer mapping. If you have addresses on two files and have geocoded the addresses, you can match records having the same address. In another Authority project, for example, we are matching homicide addresses to a list of convenience store addresses, in order to obtain a data file of convenience store homicides.

Also, you can link incident (point) data with area data through geocoding. One way to do this is to add area-level data to an incident file (for details, see *Criminal justice jurisdiction areas*, page 10). Another way to link point and area data is to add summary, aggregate incident data to an area file. Census data and many other kinds of community data -- public health data, for example -- are area-level. This means that information is available only about the area aggregate, within some set or arbitrary boundary. In order to link such area-level data to point data, we can capitalize on the fact that an area can be represented by a point (the centroid; see *Glossary*) within the area. Mapping software such as MapInfo will calculate the centroid within any area defined by a boundary (not only officially-designated areas like Census tracts or blocks, but also police districts or beats that may change over time, community or group-defined areas like street gang territories, or Hot Spot Area boundaries around the densest clusters of incidents). Incidents can be aggregated and counted within any of these defined areas, and the information then linked to the respective area-level file.

In addition, estimates of the population within a street gang territory or a Hot Spot Area can be calculated by summing the populations of all Census blocks in which the centroid is within the turf or the hot spot (see *Population information*, page 18). This is valuable for calculating rates -- the density of drug offenses within different territories, for example. See Block (1994) for details and examples.

Street gang territories:

Street gang territory maps have great potential for usefulness but can become extremely complex because of perceptual differences in defining territories, the variability of the territories over time and the level of detail that best serves problem-solving needs.

Street gang territory perceptions depend upon several things. A tactical officer may have a different view of street gang territories than a patrol officer, and a narcotics officer may see it in yet another way. The view of a community worker or neighborhood resident may differ from both. Beat assignment, which increases familiarity with an area, may also result in a different perception of street gang territories. Time of day can alter a street gang territory boundary. Night watch officers may have a different view than officers on the day watch, since different activities with different participants are likely to occur during different watches. The important thing about these differences is that they are not errors but rather information from different perspectives. If we used information from all of these perspectives, we would have a more accurate idea of street gang territories than if we used information from only one perspective.

Street gang territories also change over time -- possibly even short time periods. One purpose of turf battles is, in fact, to change turf boundaries. It is important to develop a method to update your street gang territory file. At the same time, current information is not enough. To do trend analysis or to monitor change over time, we must not destroy information about past territories. Therefore, we must maintain both current files and past files. These multiple files require careful documentation. Multiple territory maps for different perspectives and different time periods can present a formidable data management challenge.

No matter how easy the GeoArchive database is to use, complex sets of multiple territory maps may involve too much time and effort for quick decisions. Many users may need to make rapid tactical decisions that require a simple, summary territory map. In addition, a territory map may contain many levels of detail. For example, in Chicago, the Disciples street gang controls a large geographical area, but it has many factions, each controlling its own segment of that turf. Does the end user need to know the specific detail of each street gang turf, or would this be so much information that the size of the dataset would become a barrier to quick decision-making?

A GeoArchive should be developed, organized and maintained with the needs of specific groups of users in mind, such as tactical officers, patrol officers or community liaison officers. Different users have goals that require different levels of accuracy and detail. For example, the information needs of tactical and investigatory support personnel are not the same as crime analysis; the needs of short-term decision-making versus long-term planning vary; and the information necessary for developing a crime prevention strategy may not be the same as the information necessary to apprehend an offender. The GeoArchive may need to maintain multiple maps or sets of maps for each of these different kinds of users. Maintaining appropriate maps or map sets for each kind of user requires constant attention to keep territory maps from becoming obsolete. The GeoArchive developer must reconcile the need for detail -- multiple maps with different levels of information -- with the need for quick and accurate information in one "best" territory map.

The decision about what constitutes a "best" territory map depends upon local characteristics, such as the level and type of street gang activity in an area and the resources at the community's disposal to intervene in street gang activities. The need for detail must be balanced against GeoArchive management constraints and resources such as time and personnel.

Street map data:

Even for small cities, a street file is a highly complex dataset. In the United States, the usual source for street files is the Census Bureau TIGER files, available for every county. As received from the Census Bureau, TIGER files may contain many errors, such as missing or inaccurate streets or misspelled street names.[1] In addition, newly-created streets or street name changes can affect the timeliness of the street map data. The problem of error in street files affects not only the display of streets on a map but the geocoding of address data. Erroneous maps are, therefore, a serious threat to accurate decisions.

Those most likely to discover these errors are the individuals who work the streets and know them well or the analyst who analyzes the data and maps the patterns. Therefore, street files should be corrected and updated at the local level. However, local updates without central coordination means that over time, and probably over a short period of time, the city map in one station house will not be the same as that in another.

To handle this dilemma, we recommend that the GeoArchive team establish a procedure under which local-level corrections and enhancements to street files are regularly transmitted to a central "holding file." The manager of this holding file should check the accuracy and consistency of each correction and enhancement (regardless of whether it originates from a local or central source), and provide frequent, regular, periodic downloads of updated street files from the central to the local site. Until approved by this clearinghouse procedure, the updated or corrected information could be used *at the local level only*. After central clearinghouse approval, the corrections would become the standard for all dataset users.

Multiple sources of changing information present an additional problem. When data are continually accumulating, such as law enforcement information generated every minute, it may be vital that the end user has the most current information possible. However, older patterns and trends may be equally vital in order to define and solve a problem. The GeoArchive team must decide whether new data received from a source will include corrections and updates of past data. It is essential to ensure that data newly received from the source will not write over and destroy changes made to your archived data.

None of these problems can be dealt with by simple automation; they are management and coordination problems. Further, there are no guidelines provided by others who have solved the problems. For example, at this writing, the U. S. Census still has not established procedures for adopting locally-discovered corrections to the 1990 TIGER street files. In Chicago, we have been confronting these issues but have not yet solved them.[2] The most promising avenue for solution seems to be the Chicago GeoArchive Working Group, established by the Chicago Police Department and the Authority to develop procedures and methods for the coordination of data between central police department units and local users of the data.[3]

In summary, it is important to synchronize local needs and central resources. To avoid data incompatibility across local-level geoarchives and to make large and widely-used datasets (such as street files, geocoded incidents, land use or other community data) easily available at the district or neighborhood level, it is better to maintain many large, important datasets centrally. On the other hand, the best source for data validity-checking and correction is at the user level, which is often the local level. We suggest a clearinghouse procedure, designed so as to increase, not reduce, the timeliness and accuracy of data at the local level.

Summary: managing multiple databases and data sources:

The above examples illustrate the following considerations in managing multiple databases and data sources:

- **Linking databases**. Database management is critical to developing accurate, current and easy-to-use databases that have relationships to one another and that can be used together for community problem identification, development of prevention strategies, crime analysis or tactical planning. Multiple files, one for each method of linking, are necessary for problem solving. However, they generate complex data relationships requiring careful and conscientious management.

- **Rapidly changing and accumulating information**. The synthesis of information for users of varying needs requires the analysis of user needs and careful data management. How information changes also requires control. Planned procedures to update this information will dictate how changes are implemented and how previous versions of maps or other files are archived.

- **Synchronization between data source and GeoArchive.** There should be an easily accessible and quick conduit for local data users to inform a central data repository (clearinghouse) of errors and updates, and to receive updated and corrected datasets in return. Setting up such a synchronized system is not so much an automation problem as an organizational problem. The goal of Chicago's GeoArchive Working Group is to develop such a system.

This section illustrates some of the problems in the management of multiple databases and data sources issues that the GeoArchive developer confronts. Synchronization and updating of databases and files is a formidable task, as is meeting the challenge of data integrity while contending with rapid changes to the data. Databases require updating and a procedure for keeping accumulating information organized. Ensuring that local data are synchronized with the central data source is a great challenge. However, linking and synchronizing databases can provide enormous analytical power and flexibility.

Managing Data with the End User in Mind

Different sorts of problems require different kinds and levels of information. Tactical, investigatory or analytical (for example, pattern and crime analysis) decisions call for different information. For example, data required to support short-term problem-solving of an immediate threat to neighborhood safety must be current and specific to a particular situation, while data required for the analysis of long-term trends or patterns should cover a long time span and a variety of areas. Similarly, the problems of solving crime (investigation) and preventing crime, may involve profoundly different kinds of analysis and require very different kinds of information.

Datasets maintained in a central division or unit, such as police headquarters, are likely to be more stable and consistent than local data sources. Central data archives may also be more accurate, if the data manager integrates updates and corrections discovered by users, who are often at the local level. However, centrally-maintained and quality-controlled datasets may also be less timely and less relevant to immediate neighborhood concerns than current local information, such as citizen-reported concerns, tips, rumors, and observations. Local-level, current (and possibly unverified) datasets may provide a quicker response to a crisis situation. Dealing with the dichotomies between central and local information, and between verified and non-verified information, is perhaps the major challenge of GeoArchive development.

The data files in a GeoArchive consist of cases with numerous variables and a great amount of information. Some variables, such as the incident file's "offense" field with more than 400 categories of crime description, have far more detailed information than a user typically needs. Different kinds of users require different levels of accuracy and detail. It may be necessary to keep multiple street maps or street gang territory maps for each of them. In addition, these kind of data are subject to rapid change. The dual challenges of maintaining appropriate files for each kind of user and keeping them up-to-date require constant attention from the GeoArchive manager.

In this section, we describe two kinds of files: an incident file and a land use file. We show how these files can be managed by the GeoArchive developer for easy access by the end user.

Incident data:

Two methods are helpful in preparing incident data for use in analysis. The first is to create a separate database containing only those incidents of interest. For example, the GeoArchive of the Early Warning System for Street Gang Violence project contains only street gang-related incidents. The second method is to create a summary variable that consolidates codes. For example, an offender might be charged with any of several different drug offenses. For most analytical purposes, it may not be necessary to distinguish among each separate kind of drug charge. A variable that flags all drug-related incidents can speed analysis by eliminating unnecessary detail.

Selecting cases. An example of the first method can be found in the GeoArchive of the Early Warning System for Street Gang Violence project, in which a database was created containing only street gang-related crimes (see *Street gang territories*, page 30). The criterion for selecting cases as street gang-related is the presence of any one of several codes on the incident record.[4] These codes include a victim gang code, an offender gang code, a code for whether or not an automobile was used in a gang crime or for a drive-by shooting, a code for the type of gang-related activity (such as retaliation, recruitment, representing and turf) and a code for the area where gang activity took place.[5] Any of these conditions warrants inclusion as a street gang-related incident.

Recoding fields. The Early Warning System for Street Gang Violence also provides an example of the second method of preparing incident data for easier analysis -- consolidating offense codes. For analytical purposes, street gang-related incidents were classified into four types: homicide, violent, drug and other, based on the offense codes. (Most street gang-related "other" offenses are property damage with graffiti.) Analysis using this recoded variable found some street gangs (entrepreneurial gangs) were more involved with drug trafficking, and other street gangs (turf gangs) were primarily engaged in battles for territory.[6]

Land use data:

For most purposes, it would be useless to map Chicago's Sanborn file as it is (see *Data for Crime Analysis: Property information*, page 12). Since the file contains information on every piece of property in the city, a map of all the information would be blanketed with points. A thematic map (see *Glossary*), in which every address is represented by a point or icon (see *Glossary*) denoting the type of parcel, would be useful only on a very small scale, such as one block or a small group of blocks.

Such an extraordinary amount of information can be made more useful and easier for end users to access by extracting information to create new, focused, GeoArchive data files. In the Sanborn file example, abandoned buildings, schools, taverns and retail space were extracted and placed in separate point files for use as needed in analysis.

Standards and Procedures for GeoArchive Management

Each GeoArchive should have specific standards and procedures to control how data are brought into it, who uses the data and who maintains the data. Standards and procedures are vital for data integrity, disaster recovery, and data sharing or data access.

Data integrity:

Once the data verification process has been completed (see *Data Verification*, beginning on page 21), the next job of the GeoArchive manager is to ensure the data remain intact and accurate. Determine which databases and files, if any, should be marked "read only." This is especially necessary when using database management systems such as Paradox and FoxPro, which permit updating of files in "browse" mode. It is easy to accidentally erase or modify a database. Standard procedures for handling regular, periodic reception of new data from the source, and for backing-up and restoring data, will ensure any accidental modification or deletions can be recovered.

Updating data locally means making changes to databases. You will need to determine how to distinguish between completely protected databases and databases that have had additions and changes made to them on the basis of local input, such as citizen or beat officer information.

Regular updates of data, received from the source, can destroy local data. File compare programs and merge programs can assist you in making decisions about the status of a database.

Disaster recovery:

While we hope disaster never befalls you, the possibility is always present. Your hard disk could crash. There could be a fire, a plumbing problem causing water damage, a power surge, a computer virus or an accidental DOS FORMAT C: command. You should have, therefore, a disaster recovery plan that enables you to restore software and databases to your computer.[7]

A GeoArchive is particularly vulnerable to disaster because of the time-consuming computer mapping tasks -- particularly geocoding large datasets or creating boundary files (a process that may require many hours of work). Be sure, therefore, to set the automatic backup function on your computer so that it saves your work at regular intervals. A general rule to follow is to decide how valuable your time is, and backup accordingly. For example, if you are willing to lose an hour, backup every hour. Especially while editing a boundary or map file, remember to save it often.

We recommend that the GeoArchive manager encourage analysts to backup every five or ten minutes when geocoding or creating a boundary file. As a general rule, it is not until a new analyst experiences the destruction of work at least one time that the importance of backing up work becomes clear. If you have a high turnover of analysts, this could be a problem.

Periodic backup to tape by the GeoArchive manager can prevent hard feelings and possible disaster. Backup files, as well as archive files of original data, should be stored off-site or in some other well-secured place. To avoid writing over a current file, the manager should establish naming conventions using the creation date of the file.

We highly recommend purchasing and using an anti-virus protection software package. The GeoArchive manager should set up everyone's terminal so that the virus software is always active. Every time an analyst boots the machine or reads a disk, virus software should be activated. In our experience, terminals in a law enforcement environment are particularly vulnerable to viruses. In addition to virus software, if you are experiencing a repeated problem, we have found installation of a terminal lock can decrease recidivism remarkably.

Finally, a disaster recovery plan is only good if it, in fact, works. We recommend, therefore, that you simulate disaster to make sure.

Data sharing and data access:

Among the data in your GeoArchive, there will be some sensitive law enforcement and community information. The level of security required depends upon: (1) the kind of information; (2) who might use it; and, (3) the sensitivity of the information.

Verified or "official" data, like "crimes known to the police" or arrests, do not require the same level of security as unverified data, like lists of suspects, contacts or citizen tips. Official data have gone through a review process and are part of official police records. The data are standardized in format, have an identification number of some kind and have standard codes. Unverified data have not been verified through an official review process and are not public information.

The end users of data may include everyone on the community problem-solving team: community agency workers and public officials, detectives and tactical officers, neighborhood patrol officers, crime analysts and others. Under community, problem-oriented policing, information may be passed along to citizens. A different kind of information may be needed by different users and to solve different problems. However, a variety of datasets may all reside in the same GeoArchive. For example, investigative information may be stored in the same computer with the more general community crime pattern information patrol officers need for use during their watch.

Both law enforcement and community data may be sensitive. For example, in the Early Warning System for Street Gang Violence project, one agency's volunteer staff was willing to provide street gang territory information. However, they were concerned that street gang members would learn they had done so. In this case, information provided by the community could have endangered people, if street gang members had access to the information.

The issue of data access is related to data accessibility. As long as the process of data access is difficult or impossible, even for authorized users, the question of who is authorized is moot. It is when technology and database management make data available and usable that the question of access arises. One of the simplest and most productive steps you can take to facilitate ease of access to data is to standardize names of fields across databases.

Summary: standards and procedures for GeoArchive management:

In this section, we describe some standards and procedures for the GeoArchive developer to follow to ensure data integrity of the GeoArchive. We recommend that the following steps be implemented:

- Assure there are write-protected copies of important files. A "read only" file prevents users from corrupting the records within it. Clearly designate separate data files that can be manipulated by the analyst (changes to a file may include consolidation of codes or elimination of non-useful fields).

- Determine levels of security required according to the type of information concerned (community versus law enforcement, verified versus unverified, sensitive versus not sensitive, individual-level versus aggregate) and the people who might use that information. The best entity to make such determinations may be a GeoArchive Working Group (see *Glossary*).

- Define clearly who may have the use of a specific database, because the information that it contains is public information. Alternatively, make it impossible for all but specified users to access a database that may contain sensitive information such as corrections release information or investigative information.

- Have a disaster recovery plan. The plan should include the following processes: (1) backup all databases and software that can be accessed for restoration; (2) program the automatic backup function on your computer to save your work (for example, geocoding data or creating a boundary file) regularly; and, (3) purchase an anti-virus protection program and set it up on your computers so the virus software is always active.

Notes

[1]See Richard Block (1994) for discussions of errors in the Chicago TIGER files.

[2]Another model for the coordination of geographic information systems is the system in Baltimore County, Maryland. For more information, contact Phil Canter of the Baltimore County Police Department.

[3]For additional information about the GeoArchive Working Group, contact Director Gregory Kowalec, Chicago Police Department Data Division.

[4]In Chicago, no single variable or code marks a case or an incident as street gang-related.

[5]For more detail about these codes, see the *Early Warning System GeoArchive Codebook* (Green & Whitaker, 1994).

[6]For more detail, see Block and Block (1993).

[7]For more information about disaster recovery, see ICJIA (1989; Section III: Security). Also see *Mislabeled Files or Databases* elsewhere in this report (beginning on page 21).

Appendix I: Advisory Board Members: Early Warning System Project[1]

Commander John Kozaritz
Area Four Detective Division
Chicago Police Department
3151 West Harrison St.
Chicago, IL 60612

Sgt. Ronald Rewers
Case Management Sergeant
Area Four Detective Division
Chicago Police Department
3151 West Harrison St.
Chicago, IL 60612

Barbara McDonald, Director
Research & Development Division
Chicago Police Department
1121 South State St., Room 401
Chicago, IL 60605

Lt. Phillip Chomiak
Research & Development Division
Chicago Police Department
1121 South State St., Room 401
Chicago, IL 60605

Lt. John Kennedy
Administrative Officer
Special Functions Group
Chicago Police Department
1121 South State St., Room 1223
Chicago, IL 60605

Glenda Cunningham
Neighborhood Development Organizer
Bethel New Life, Inc.
4746 West Rice St.
Chicago, IL 60651

Robbie Jacquette
Consultant
P.O. Box 12204
Chicago, IL 60612

Joseph Guerrero
Park Supervisor, Dvorak Park
1119 West Cullerton St.
Chicago, IL 60608

Anthony Mata
Research Analyst
IL Criminal Jus. Information Authority
120 South Riverside Plaza
Chicago, IL 60606

Sonia Gaete
2938 West 25th Place
Chicago, IL 60623

Ann Cibulskis, Senior Planner
Research & Program Development Dept.
Chicago Housing Authority
22 West Madison St.
Chicago, IL 60602

Taya D. Sun, Director
Preventive Programs
Chicago Housing Authority
22 West Madison St.
Chicago, IL 60602

Rey Colon
Program/Volunteer Coordinator
Boys & Girls Clubs of Chicago
Corporate Headquarters
625 West Jackson Blvd.
Chicago, IL 60661

Christine Corrado, Director
Training and Program Development
Boys & Girls Clubs of Chicago
625 W. Jackson Blvd.
Chicago, IL 60661

Professor Richard L. Block
Loyola University of Chicago
Department of Sociology
6525 North Sheridan Rd.
Chicago, IL 60626

Professor Michael Maltz
Department of Criminal Justice
University of Illinois at Chicago
Box 4348, M/C 141
Chicago, IL 60680

Professor Irving Spergel
Project Director
Gang Violence Reduction Project
School of Social Service Administration
969 East Sixtieth St.
Chicago, IL 60637

Ex-Officio Board Members

Deputy Chief James A. Maurer
Detective Division
Chicago Police Department
1121 South State St.
Chicago, Illinois 60605

Chief Joseph Beazley
Joliet Police Department
150 West Jefferson St.
Joliet, Illinois 60431

Officer Art Block
Special Operations
Chicago Police Department
3151 W. Harrison St.
Chicago, IL 60605

Sherwood S. Williams
Chief of the Facility
Markham Court House
16510 South Kedzie Ave.
Markham, IL 60426

Robert W. Dart
General Manager, Security Services
Chicago Transit Authority
Merchandise Mart
PO Box 355, Room 705
Chicago, IL 60654

Note

[1]This list represents the membership of the Early Warning System Project Advisory Board, as of May 1994.

Appendix II: A Brief Overview of the Early Warning System for Street Gang Violence[1]

In Chicago, as in many other parts of the country, homicide is a pervasive and increasingly serious problem. The homicide situation becomes even more discouraging if we agree with the conventional wisdom that says there is nothing we can do about it: that homicide cannot be prevented. However, as Chicago Police Department Superintendent Matt Rodriguez was quoted as saying (*Chicago Sun Times*, September 9, 1992, page 1), "It's something we're not proud of, but it's something we've got to do something about."

Fortunately, a growing body of research indicates that the conventional wisdom is not true. In fact, *many homicides can be prevented.* To prevent homicides we must be aware of three things:

- **First**, we must identify the specific type of homicide. Intervention strategies that work for domestic violence will not work for street gang violence. Strategies to prevent robbery homicides may be ineffective for barroom brawls.
- **Second**, to prevent homicide, it's not enough just to look at homicide incidents. You must also look at all the events that may have led up to the violence, whether they were nonlethal criminal offenses like graffiti and vandalism or narcotics activity or information about the community and the neighborhood situation.
- **Third,** just identifying the problem is not enough. We must use the past to predict the future. In other words, we must summarize this vast amount of data and come up with specific information the community and the police can use for intervention strategies.

The Early Warning System project, which is now underway in Police Area Four, an area containing some of the riskiest neighborhoods for street gang violence in Chicago, is attempting to do something about this homicide epidemic. The project uses computer mapping, the Illinois Criminal Justice Information Authority's STAC package and community and police data organized in a GeoArchive management system to identify specific neighborhoods at high risk for a surge in serious street gang violence and homicide while there is still time to intervene and save lives.

The specific purpose of the project is to develop an automated early warning system for law enforcement that will identify potential neighborhood crisis areas, areas that are at high risk for suffering a "spurt" of serious street gang-related violence and homicide. This early warning system is based on a statistical model that consolidates spatial information obtained from a variety of community and law enforcement sources and uses automated Hot Spot Area identification and other geographic statistics as tools to target crisis neighborhoods.

In any year, homicides in Chicago practically cover the city's map. Yet if you look closely, there are patterns. And these patterns are different for different types of homicide. But how do we make sense of the clusters of street gang homicides? We can use technology -- the Hot Spot Area program in STAC. Hot Spot Area analysis summarizes vast amounts of information and bases these summaries on the real events themselves, not artificial boundaries.

To be successful in reducing levels of death and injury from violence, we must develop intervention strategies based on accurate, timely and complete information about specific incidents that are specific types of violence, occurring in defined areas of the city and affecting a targeted group of people. First, we must identify the problem; second, we must focus intervention strategies on that specific problem. The Early Warning System project in Chicago Police Area Four is doing this. It links computerized mapping technology and STAC to focus intervention strategies on high-risk neighborhoods and highly vulnerable people -- using *high tech* and *high touch* methods to prevent homicide.

STAC and the Early Warning System can be seen as the **information foundation for community policing**. In the Early Warning System we are demonstrating it is possible to prevent violence if we draw on community and law enforcement data sources, use statistical tools like STAC to identify the high-risk areas and situations and focus intervention strategies on the targeted problems. Violence reduction is a three-step process:

- Identify the specific problem (using community and law enforcement information)
- Target *potential* crisis areas
- Use community and law enforcement resources to intervene

But STAC has additional benefits beyond its statistical power. STAC and the GeoArchive database provide an automated institutional memory, allow the combined analysis of data from law enforcement and community sources, increase the cooperative sharing of information within the police department and between law enforcement and the community and produce results that can be communicated in non-technical terms. Most important, the project has initiated a revolutionary change in Chicago policing. Information and the capability to analyze that information are no longer monopolized by central headquarters but are instead available to officers and community members working to reduce crime in the neighborhoods.

Note

[1]Excerpted from Block (1992). See also Carolyn Rebecca Block (1991).

Bibliography

Benzon, Bill, Rick Frei, Lorene Doyle and Janet Brown (1991). *MapInfo for DOS: User's Guide, Version 5.0.* Troy, New York: MapInfo Corporation.

Block, Carolyn Rebecca (1984). How to Handle Seasonality: A User's Guide to the Detection and Analysis of Seasonal Fluctuation in Criminal Justice Data. Chicago: Illinois Criminal Justice Information Authority.

_____ (1987). Is Crime Predictable? A Test of Methodology for Forecasting Criminal Offenses. Chicago: Illinois Criminal Justice Information Authority.

_____ (1992). Automated spatial analysis as a tool in violence reduction. Paper presented at the San Diego Problem-Oriented Policing Conference, Police Executive Research Forum and San Diego Police Department.

_____ (1991). Early Warning System for Street Gang Violence Crisis Areas: Automated Hot Spot Identification in Law Enforcement. Chicago, Illinois: proposal to the Bureau of Justice Statistics.

_____ (1994). STAC Hot Spot Areas: A statistical tool for law enforcement decisions. *Proceedings of the Workshop on Crime Analysis through Computer Mapping.* Carolyn Rebecca Block and Margaret Dabdoub (Eds.). Chicago: Illinois Criminal Justice Information Authority.

_____ and Richard Block (1993). *Street gang crime in Chicago.* Research in Brief, National Institute of Justice. Washington, D.C.: U.S. Department of Justice.

Block, Richard (1994). Practical ideas on using, correcting and updating TIGER files: Chicago. *Proceedings of the Workshop on Crime Analysis through Computer Mapping.* Carolyn Rebecca Block and Margaret Dabdoub (Eds.). Chicago, Illinois: Illinois Criminal Justice Information Authority.

_____ (1977). *Violent Crime: Environment, Interaction and Death.* Lexington, Massachusetts: Lexington-Heath.

Bobrowski, Lawrence J. (1988). Collecting, organizing and reporting street gang crime. Presented at the American Society of Criminology meetings. Chicago Police Department, Special Functions Group.

Dubrow, Nancy F. and James Garbarino (1989). Living in the war zone: Mothers and young children in a public housing development. *Child Welfare*, 68 (1):3-15.

Georges-Abeyie, Daniel E. (1981). Studying black crime: A realistic approach. Pp. 97-109 in Pat Brantingham and Paul Brantingham (Eds.), *Environmental Criminology*. Beverly Hills: Sage.

Green, Lynn A. and Robert Whitaker (1994). *Early Warning System GeoArchive Codebook: Area Four Project*. Chicago: Illinois Criminal Justice Information Authority.

Illinois Criminal Justice Information Authority (1991). *Trends and Issues 91: Education and Criminal Justice in Illinois*. Chicago: Illinois Criminal Justice Information Authority.

_____ (1989). *Annual Audit Report for 1988: Illinois' Computerized Criminal History System*. Chicago: Illinois Criminal Justice Information Authority.

Kuhn, Thomas (1977). *The Structure of Scientific Revolutions*. Chicago, Illinois: University of Chicago Press.

LeBeau, James L. (1987). The methods and measures of centrography and the spatial dynamics of rape. *Journal of Quantitative Criminology* 2 (2,June):125-141.

Maltz, Michael D., Andrew C. Gordon, Robert K. Lebailly, Warren Friedman, Marc Buslik and Matthias Casey (1987). Crime Statistics in the Small: Mapping Individual Events. Presented at the annual meeting of the American Society of Criminology.

_____ , Andrew C. Gordon and Warren Friedman (1990). *Mapping Crime in its Community Setting: Event Geography Analysis*. New York: Springer-Verlag.

Rengert, George (1994). Comparing cognitive hot spots to crime hot spots. *Proceedings of the Workshop on Crime Analysis through Computer Mapping*. Carolyn Rebecca Block and Margaret Dabdoub (Eds.). Chicago: Illinois Criminal Justice Information Authority.

Roncek, Dennis W. (1992). Housing projects and crimes revisited. Paper presented at the 1992 annual meetings of the American Society of Criminology.

______ and Ralph Bell (1981). Bars, blocks, and crimes. *Journal of Environmental Systems*, 11:35-47.

_____ , Ralph Bell and Jeffrey M.A. Francik (1981). Housing projects and crime: Testing a proximity hypothesis. *Social Problems*, 29: 151-166.

_____ and Richard Block (1983). The effect of neighborhood change on homicide in Chicago. Paper presented at the annual meetings of the American Society of Criminology.

_____ and Donald Faggiani (1985). High schools and crime: A replication. The *Sociological Quarterly*, 26(6):491-505.

_____ and Antoinette Lobosco (1983). The effect of high schools on crime in their neighborhoods. Social Science Quarterly, 64(3, September):598-613.

_____ and Pamela A. Maier (1991). Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of "hot spots." *Criminology,* 29(4):725-753.

Smith, Ben and Howard Eglowstein (1993). Solutions focus/desktop mapping software: Putting your data on the map. *Byte*. January:188-200.

Spergel, Irving A. (1984). Violent gangs in Chicago: In search of social policy. *Social Service Review* 58(2, June):199-226.

Stephenson, Larry K. (1980). Centrographic analysis of crime. Pp. 247-263 in Daniel E. Georges-Abeyie and Keith D. Harries (Eds.), *Crime: A Spatial Perspective.* New York: Columbia University Press.

Glossary

Area data: Information about a geographical area, such as a Census area (block, tract), a police district or a street gang territory. Area information may be linked to a boundary around an area, or it may be linked to a central point or intersection within the area (see *centroid*).

Boundary data: A dataset for computer mapping that consists of lines (sometimes called polygons), each of which encloses an area.

Centroid: "A centrally located point within a boundary" (Benzon *et al.*, 1991).

Database files (Data tables): Sets of information including an identifier, organized in a file structure (see *file structure, fields*).

Digitize: The process of translating map features (such as streets, expressways, or bodies of water) to an x- and y- coordinate grid for computer mapping (see *map data*).

Early Warning System for Street Gang Violence Crisis Areas (Early Warning System): A project that uses a GeoArchive and the Spatial and Temporal Analysis of Crime (STAC) software to identify neighborhoods at risk for a crisis of street gang violence (see Appendix II).

Fields (variables): Columns in a database. Each field contains a type of information, such as type of offense, date, address or offender's street gang (see *database*).

File structure: A map or chart showing the location of the fields on a database (see *fields*, *database*).

GeoArchive: An information foundation for community policing: a GIS that contains address-based data from law enforcement and community sources and is set up so that it can be updated, maintained, mapped and analyzed at the neighborhood level (see *GIS*).

GeoArchive Working Group: Representatives from central and local GeoArchive users who work together to develop procedures for managing, developing and maintaining the GeoArchive.

Geocode (geocoding): (1) the assignment of x- and y- coordinates, such as longitude and latitude, to an address, so that the address can be mapped; (2) the x- and y-coordinates that correspond to a given address; (3) the assignment of an event, incident or map feature to an area, such as a Census tract or a police district.

Geographic Information System (GIS): A database containing a digitized map and data that is geocoded to be located on that map (see *GeoArchive, database, geocode, map data*).

Hot Spot Area(s): The densest concentration(s) of points on a map. The Hot Spot Area search routine in the STAC package searches for the densest cluster(s) of points regardless of arbitrary boundaries and defines them by a standard deviational ellipse. Hot Spot Areas turn point data into area data (see *STAC, standard deviational ellipse*).

Icon: A symbol on a map, representing a point. For example, O'Hare Airport might be represented by a small plane; homicides might be squares while other violent incidents are circles (see *thematic map*).

Map data: A dataset containing digitized information on streets and their address ranges for a geographical area, used both for displaying streets and for geocoding addresses into point data (see *GIS, geocode, digitize, point data*).

Point data: Address-based data containing information about geographical points, for example, the location of crime incidents or the location of businesses that hold liquor licenses or data containing information about an area, such as the centroid of an area (see *centroid, geocode, GIS*).

Spatial and Temporal Analysis of Crime (STAC): STAC is a software package developed by the Authority, which contains a toolbox of spatial analysis statistics for analyzing the distribution of events on a map. (It is not a mapping package.) STAC is available from the Illinois Criminal Justice Information Authority at no cost to law enforcement agencies.

Standard deviational ellipse: "By rotating the original reference axes (from which the measurements for mean center and standard distance were made), it is possible to establish a new set of axes along one of which the variance of the X_is or Y_is is maximized and along the other the variance is maximized" (Stephenson, 1980: 151). Also see LeBeau (1987) (see *Hot Spot Area, STAC*).

Thematic map: A map displaying "some other data theme besides geographic location" (Benzon *et al.*, 1991), such as street gang territories, population information or type of offense. With a point or pin map, each different kind of point (for example, violent incidents versus property incidents) is represented by a different icon. With area data, each area's categorization on some characteristic (such as the street gang territory or the level of crime in the area) is represented by shading over the area or by an icon placed at a central point within the area. (see *icon, point data, area data*).

TIGER file: Digitized street map data created by the U. S. Census Bureau for the 1990 Census, for all the streets in every county in the United States. Available from the U.S. Census Bureau. For more detail, see Green and Whitaker (1994) (see *map data, digitize*).

Index

ILLINOIS
CRIMINAL JUSTICE
INFORMATION AUTHORITY

120 South Riverside Plaza Chicago, Illinois 60606 (312) 793-8550