Developing GIS-Enabled Crime Analysis Applications Using Value-Added Warehouse Data for Law Enforcement

Mark Dougherty
Information Systems Director
St. Louis County Police Department
7900 Forsyth Blvd, St. Louis, MO 63105
(314) 615-7826
mdougherty@stlouisco.com

Law Enforcement Systems Manager
Regional Justice Information Services (REJIS)
4255 West Pine Blvd. St. Louis, MO 63108
(314) 535-1455 ext. 292
mdougherty@rejis.org

Hsiu-Hua Liao, Ph.D.
Senior Application Programmer
St. Louis County Police Department
7900 Forsyth Blvd., St. Louis, MO 63105
(314) 615-7830
hliao@stlouisco.com

Paul Trudt
GIS Analyst
St. Louis County Police Department
7900 Forsyth Blvd., St. Louis, MO 63105
(314) 615-5337
ptrudt@stlouisco.com

Developing GIS-Enabled Crime Analysis Applications Using Value-Added Warehouse Data for Law Enforcement

Mark Dougherty, Hsiu-Hua Liao, Paul Trudt
St. Louis County, Police Department

ABSTRACT

Information sharing among law enforcement agencies has become increasingly important in recent years. In the St. Louis area, a regional data warehouse system called Crime**MATRIX** (**M**ulti-jurisdictional **A**naly**T**ical **R**epository for **I**nformation e**X**ploitation), has been established for data sharing by the St. Louis County Police Department (STLCPD) and the Regional Justice Information Service (REJIS). It is actively serving over 40 local, county and federal law enforcement agencies in the St. Louis area. Two different yet integrated applications, RAMS and LYNX, were developed to access the value-added data in the new St. Louis regional CrimeMATRIX data warehouse. RAMS is a desktop application that allows users to map crime events, apply spatial filters, and create a short-list of potential suspects based on the information given about crime events, locations and activities. LYNX is a web-based application that provides detailed summary data about a "person of interest" based upon data originating from multiple and dissimilar information system platforms. Both applications are geo-enabled and utilize mapping components to provide the spatial picture of persons and/or events.

INTRODUCTION

The developments in Geographic Information System (GIS) technology have come a long way in the past two decades. Since the implementation of the Canadian GIS in 1964, the field and application areas of GIS have grown rapidly, creating an enormous literature explosion in its wake and generating massive interests worldwide. The domain of current application areas of GIS include: environmental planning; natural resource conservation; health care and emergency planning; transportation and utilities management; agriculture and forestry; coastal zone planning and real estate management. In fact, GIS has been used in any field for which the handling, manipulation, and analysis of spatially referenced data is part of the analysis and decision-making process.

In late 1990s, GIS technology has gradually been implemented to identify hot spots, and analyze the spatial patterns of crime within the law enforcement community. During this time, the St. Louis County Police Department (STLCPD) acquired the ArcView GIS mapping software from the Environmental Systems Research Institute (ESRI) for crime analysis and planning applications. However, the effort was unsuccessful due to the following problems. First, the quality of the address data captured in the Records Management System (RMS) and Computer Aided Dispatch (CAD) systems were poor. Also the Tiger/Line street data that was available at the time for geocoding was poor. Both factors contributed to a 46% match rate for plotting crimes and calls for services. Secondly, data had to be exported from the RMS and CAD systems and restructured to fit the geocoding format. No address cleansing routine was used to improve the match rate, etc.

It wasn't until the Police Department needed to replace its legacy CAD System in 1999, because of Y2K incompatibility problems, that GIS integration took a great leap forward. The CAD System had a requirement that all call for service data had to be geocoded by the CAD mapping software prior to dispatching. This forced the Department to invest a significant amount of funds and GIS programming resources to create a centerline map for the County's 526 square mile roadway system. Besides supporting the CAD system, the new map provided other benefits. It provided the GIS infrastructure to develop Automated Vehicle Location (AVL) software and Crime mapping applications. The development of the Department's data warehouse and crime mapping applications were a result of this CAD GIS development effort.

DATA WAREHOUSE FOR LAW ENFORCEMENT

CrimeMATRIX

The CrimeMATRIX is a law enforcement data warehouse system that takes a regional approach to the collection, standardization, association analysis, and sharing of criminal justice information among local, state and federal agencies. The system is an Investigative Support System (ISS) for law enforcement. The system allows agencies to search information pertaining to people, vehicles, crimes, telephone numbers and property across jurisdictional lines. Corresponding relationships between subject areas such as persons having the same telephone numbers are revealed by means of a linking analysis system. This data warehousing system allows the investigator the ability to view people, places and things independently and also as entities sharing common data elements. In addition to querying its own data repository, the system makes real-time interface calls to State driver license and State and National Criminal Information Center (NCIC) wanted files for a more national approach to identifying a person of interest. Figure 1 shows the overview of the CrimeMATRIX data warehouse.

Data Collection

The CrimeMATRIX harvests data from a number of local law enforcement systems. In the St. Louis area, the system harvests mug shot and sex offender records from a regional system called IRIS (Image Resource and Imaging System). This data is harvested on an hourly basis. Police report information from a regional record management system CARE (Computer Assisted Report Entry) is harvested daily. Traffic ticket data is captured daily. Gang and Gun Permit data are harvested on a bi-weekly basis. Probation, parole and death certificate information are harvested from state systems on a monthly basis. In each of these systems, pedigree, address, vehicle, telephone numbers and crime information are all extracted and imported into the CrimeMATRIX data repository. Table 1 shows the description of each data source and the system specification.

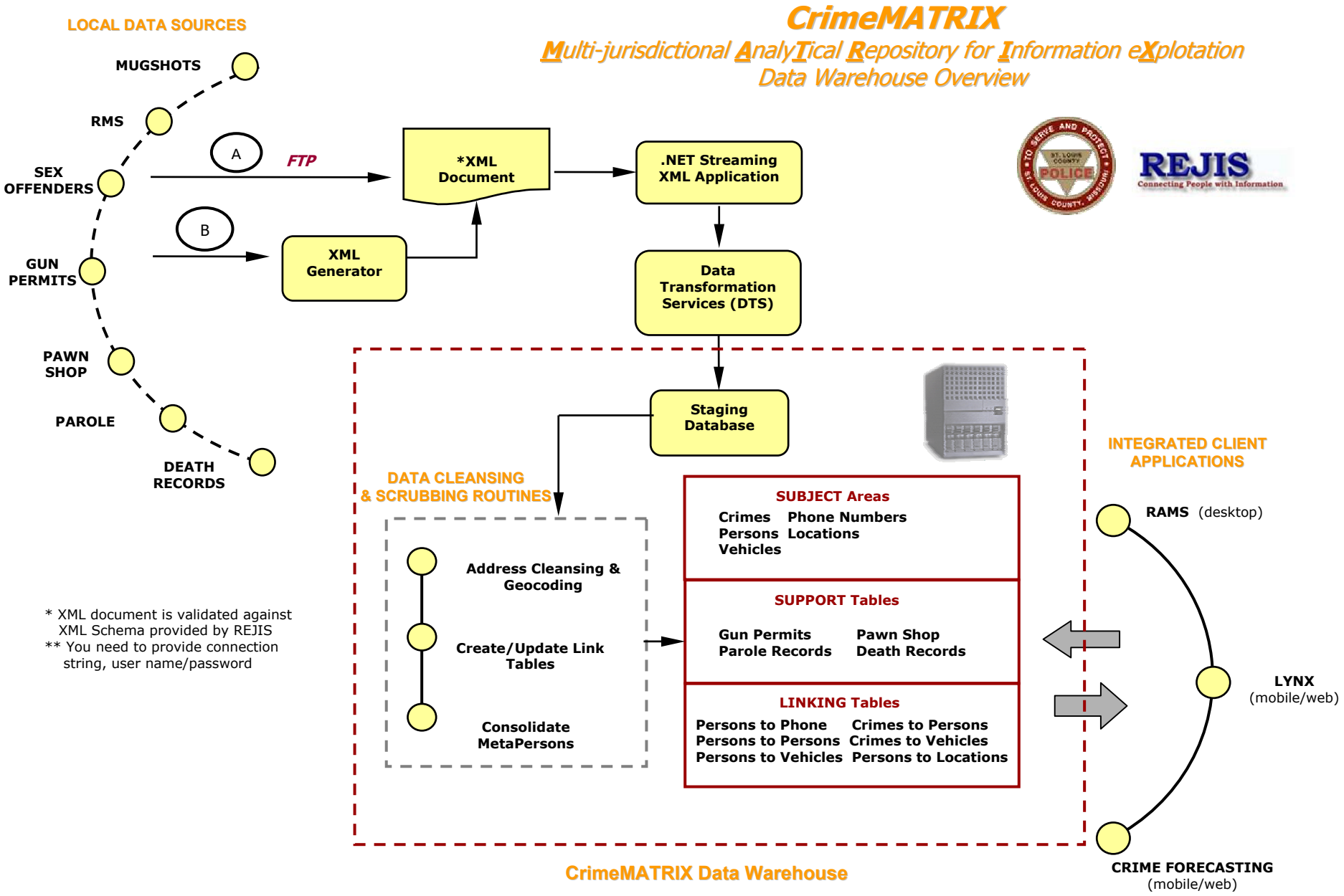

Figure 1. CrimeMATRIX data warehouse overview

Table 1. Data Sources and System Specification

Data Sources	Platforms	Description
CAD	DB2/Unix	E911 Dispatch System Call For Service
CARE	ORACLE/Intel	Record Management System
IRIS	MS SQL Server 2000/Intel	Mugshot and Sex Offender Booking
TICKET	MS SQL Server 2000/Intel	Traffic Violation
Pawnshop	MS SQL Server 2000/Intel	Pawnshop monitoring
Gun Permit	MS Access/ORACLE/Intet	Gun Permit Information
GIANT	MS Access/Intel	Gang Information Tracking
Probation & Parole	AS400	Obtained monthly from Missouri State
Death Files	AS400	Obtained monthly from Missouri State

Data Standardization

Because the CrimeMATRIX harvests data from a number of different information systems, the manner in which each application stores its data varies considerably. In the majority of cases, these applications do not share a common data dictionary or data formatting convention. As shown in Table 2, a phone number can be stored as a string with or without hyphens (xxx-xxx-xxxx) or as an integer (xxxxxxxxxx). In addition, these information systems reside on different operating system platforms such as Unix, Mainframe or Microsoft and utilize different relational database structures such as DB2, Oracle, MS SQL and MS Access, or they utilize other hierarchical data structures.

Table 2. Various Data Format for Phone in Different System.

System	Field Name	Field Type	Field Format	Field Length
CARE	szhomephone	String	xxxxxxxxxx	10
IRIS	res_phone	String	(xxx)xxx-xxxx	13
Probation & Parole	Addressee phone 2 area code	Numeric	xxx	3
	Addressee phone 2 exchange	Numeric	xxx	3
	Addressee phone 2 last 4 digits	Numeric	xxxx	4
Wanted Files	szphonenum	String	xxx-xxx-xxxx	10

Association Analysis

A major strength of the CrimeMATRIX system is its ability to reveal hidden relationships between individuals. These associations are exposed by means of matching common data elements in different types of records. In most instances, there are common data elements such as address, telephone, and location data that are common among varying applications. Residential and employment address information, vehicle data, home and work telephone numbers, and shared crime data are all used by the CrimeMATRIX to build linking or relationship tables. These secondary data structures enable the system to quickly establish relationships between people, places and things and process user requests for information. As persons of interest are harvested into the system they are compared and consolidated into the existing database population. The system thereby creates a meta-person record. This parent record contains the latest pedigree, address, phone number and other descriptive information of a person. It links back to its originating or child records by means of a common record identifier.

Investigative Support System (ISS) Architecture

With data stored in various systems, users are unable to perform queries on all systems simultaneously. By design, these systems are predominately transaction processing systems (TPS) (Muhammad, 1999). These systems are designed to streamline the data entry component of capturing crime data. They are used to capture data about calls for police service, crimes and offenders. Due to this focused approach, these systems generally have limited interfaces to other law enforcement applications within the agency. Interfaces to outside agency systems are generally very limited or non-existent. The reporting features of these systems consist of standard summary reports with limited ad hoc reporting tools. Due to these limitations, the systems tend to deliver a lot of data in fixed or static printed reports. For example, in a case study, it takes approximately 22 hours to gather all information from CARE, IRIS, GIANT, AFIS, and

Ticket systems on a particular suspect. It would also take an unspecified amount of time to search through all reports manually to find any links pertaining to the suspect.

Unlike transaction-based systems that are designed specifically for the rapid input and processing of data such as a CAD (Computer Aided Dispatch) system, the CrimeMATRIX system is designed as an Investigative Support System (ISS). The system is designed for providing output. It is designed to provide quick response to questions regarding crime and those who commit offenses. The system is similar to Decision Support Systems (DSS) found in the business environment (Power, 2002). Whereas, the business DSS model summarizes financial and operational data in a business environment, the ISS model standardizes and summaries law enforcement operational data such as pedigree, crime, property, address and telephone number information for investigative research. It is used by law enforcement personnel to determine crime trends and locate offenders. The system utilizes On-Line Analytical Processing (OLAP) to permit a multi-user regional law enforcement environment to query diverse data stored in a multi-dimensional data structure. In summary, the CrimeMATRIX's data structure is designed to optimize rapid ad-hoc information retrieval.

State and Federal Interfaces

An important component of a law enforcement information system solution is the ability to query both inside and outside of its environment. The CrimeMATRIX system allows the investigator to simultaneously query information from both its own database and outside state and national systems. Wanted and driver license data is obtained through this process. The CrimeMATRIX applications accomplish this feature by placing an internal application call to the State system. This call passes the person of interest's pedigree information to an existing transaction that resides on the State or Federal system. The information is returned to the investigator in a separate data window.

Regional Crime Data Sharing

One of the main objectives of the CrimeMATRIX system is to create a regional information-sharing environment. The CrimeMATRIX application allows all participating agencies to view crime data and offenders. Although, police reports in their entirety are not displayed, elements of the crime and "persons of interest" are shared. This feature allows each jurisdiction local control of sensitive case information. Narrative portions of police reports are not shared at this time. All participating agencies sign an inter-agency information sharing agreement. Because the system allows neighboring jurisdictions to view each other's crime activity, both police administrators and investigators realize a regional perspective of crime in a real-time environment. It permits agencies to coordinate proactive crime reduction programs, thereby more efficiently utilizing their department resources.

INTEGRATING GIS CRIME ANALYSIS APPLICATIONS

As discussed earlier, one component of the CrimeMATRIX system is its hybrid database structure that harvests and manipulates crime data in a near real-time environment. However, client-side applications were needed to extract the information for investigators. Currently, four applications have been written to query and process data from the system. The first application is RAMS (Report Analysis and Mapping System). This crime mapping application displays crime trends based upon specific search criteria. It also identifies possible offenders based upon their prior arrest history and the activity space (home, work, crime locations) in which they operate. The second application is called LYNX. This application allows an investigator to identify and locate offenders. It provides mug shot images, mug shot line-ups, offender mapping, criminal history, address history, probation and parole status, gun permit issuance data, pawnshop activity monitoring, and inter-relationship analysis. The third application is called Pawnshop Monitoring. This application captures people and property data relating to the pawning of merchandise. Individuals pawning property are linked to existing persons of interest in the CrimeMATRIX to create a more comprehensive view of the person. Property that is pawned is compared to stolen property records in the database. The fourth application is under development and is called MapMatrix. This application allows the investigator to utilize a web-client to map sex offenders, probation and parole offenders, and crimes in a multi-jurisdictional environment.

RAMS

The RAMS (Records Analysis and Mapping System) application of the CrimeMATRIX system is a multi-jurisdictional information management tool that plots user specified crime locations on an interactive, intelligent GIS map. The application interfaces with the CrimeMATRIX database engine to provide near real-time (less than 24 hours old) crime data for both approved and unapproved police reports. Police administrators, precinct or district commanders, neighborhood policing officers and investigators use standard UCR (Uniform Crime Report) crime categories (Burglary, Arson, etc.) and secondary crime modifiers (robbery premise, larceny nature, and property type) to further refine their search.

The system can be used to search for crime trends within a police jurisdiction or across more than 32 law enforcement agencies participating in the crime data sharing initiative. All participating police agencies sign an agreement to share certain crime elements contained in police reports and mug shot records. Thereby enabling the system to provide a multi-jurisdictional view of crime activity. This feature allows neighboring departments to cooperate and leverage resources to combat specific crime problems. RAMS can reveal crime patterns by crime type, time of activity and/or by area across multiple jurisdictions. This allows law enforcement to take a proactive approach to neighborhood policing. RAMS have been successfully deployed in the St. Louis region and actively supports the citizens of over 32 municipal law enforcement agencies. Local offices of the F.B.I., D.E.A., A.T.F. and the U.S. Attorney's Office also utilize the system.

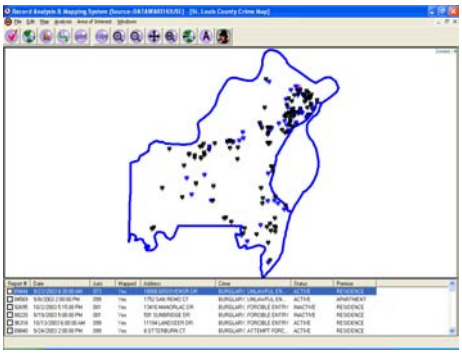

Figure 2. 4th Qtr residential burglaries w/ jewelry stolen

Because RAMS utilizes the CrimeMATRIX database engine, it is able to plot hundreds of crimes in minutes (Figure 2). The reason the application can rapidly geocode events is that the CrimeMATRIX database had performed a preliminary geocode of all crime events when they were originally harvested from the legacy systems. The x and y coordinates of all address data are stored in the database as attribute data. In addition to mapping crime events, the RAMS application summarizes thousands of crime records by day of week, time of day, crime status and geographical area or beat. In addition to the basic charts for premise type, frequency by time of day, frequency by day of week and crime distribution by precinct and/or district, the application provides a "Drill Down Utility". This utility summarizes the result of a crime query into a multi-dimensional data cube (Han and Kamber, 2001). It allows the user to dynamically dissect crime by time and location. For example, a search for all residential burglaries would identify the number of only Active cases in Precinct 1 that occurred on a Tuesday between the hours of 12:00 PM and 1:00 PM.

RAMS is a visual basic application that utilizes an enhanced ESRI Map Objects (ESRI, 1992) control for plotting crime and offenders. In addition to displaying street and jurisdictional boundary lines, the system incorporates aerial map layers and real estate property data. All of the attribute data associated with a map layer can be identified. For example, parcel layers identify the property owner, address and whether the property is owned or rented. The aerial image layer provides an additional spatial dimension to an investigation (Figure 3).

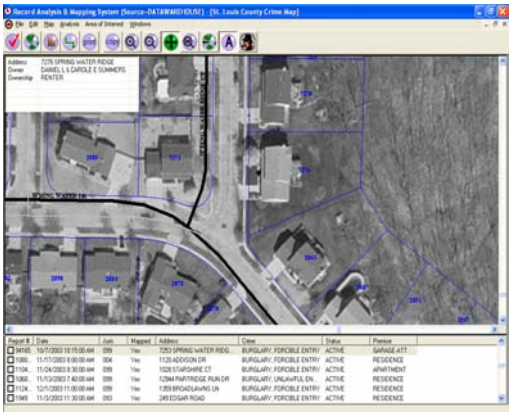

Figure 3. Residential burglary backing up to common

The aerial image layer allows the investigator to view a crime scene from an environmental criminology perspective (Rossmo, 2000). For example, an investigator may be unaware that a set of burglaries adjoins common ground areas until they activate this layer. This aerial view reveals how the burglar used the same method of access, wooded common ground area, to area homes.

In addition to mapping crime and creating a "hot spot/activity area", RAMS has the ability to generate a collection of potential suspects based upon their prior activity in a specified area. This "Generate Persons of Interest" feature prompts the user to indicate whether the POI lives, works, or has prior criminal history (active in activity area). Additional dialog search criteria are provided to the user to further refine their search by pedigree and/or specific police reports feature in RAMS can be used to generate a "spatial" list of potential suspects. RAMS utilizes pedigree, address and charge history data from the CrimeMATRIX database to create this pool of suspects. For example, if a neighborhood experiences a sudden increase in residential burglaries, RAMS would be able to show if a recent parolee was now living in the area or show locations in the neighborhood associated with persons that have a history of residential burglary (Figure 4). If a suspect's physical description is available, this data can be incorporated into the search criteria for suspects.

The RAMS application integrates the LYNX analysis application into its POI identification feature. LYNX identifies individuals plotted as potential suspects through an application interface. For example, an investigator could select an individual plotted as a known burglar and access their entire criminal history, mug shot history and known associates. In the example below, the investigator searched for all "persons of interest" who lived, worked or were arrested in the defined activity space who had a prior arrest for burglary. The red highlighted POI is further identified through an interface to the LYNX system (Figure 4). Table 3 summarize the application and time saving features for RAMS.

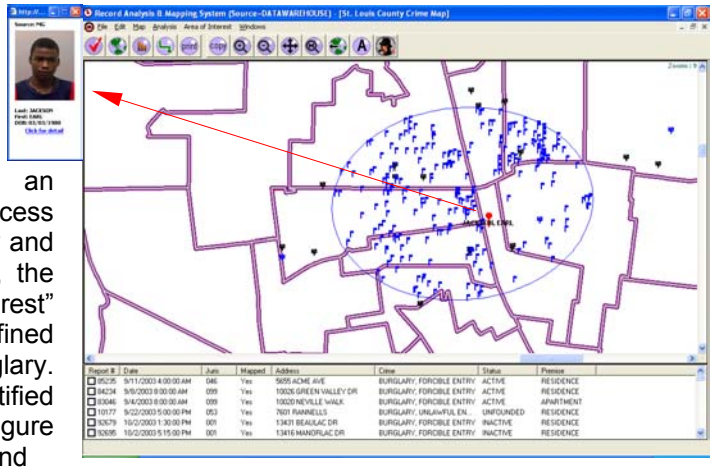

Figure 4. All POI with burglary history were active in the defined active space.

Table 3. RAMS Features

Application Features

- Automated crime mapping and analysis reports
- Maps all crime events
- Displays aerial images if available
- Allows for regional customization of map layers
- Maps Activity space of possible suspects
- Custom charts to compare time frames (MTD, YTD, etc.)
- Seamless integration with LYNX application
- Multi-jurisdictional view of crime events/patterns for any time period
- Access data from multiple data sources through CrimeMATRIX database
- Map locations of specific crimes by complaint number
- Allows for secure access and distribution of data
- Can save mapped events, images and x & y coordinates to local file for use in other applications (Arcview, etc.)
- Scalable application usage from big picture (Police Chief, Precinct / District Captain) to granular analysis (Neighborhood Policing Officer/Investigator)

Time Savings Features

- Reduces man hours while increasing efficiency
- Reduces time dedicated to 'paper chase' of searching many different systems
- Aggregates hundreds of reports in seconds by
  Day of week, Hour of day, Location type, Area, Precinct, District, Beat

LYNX

LYNX is a browser-based web application that allows an investigator to quickly identify people, places and events that are incorporated in the CrimeMATRIX system. The investigator identifies the subject area (persons, phone numbers, location, vehicles, or tickets) to search by selecting the appropriate tab. A Person Search will return a collection of people matching the search criteria. The CrimeMATRIX consolidates the many individual records (arrest, suspect, ticket, gang, etc) comprising a person, into one "MetaPerson" record. When an investigator performs a search, the resulting response page displays Meta-Persons folders (Figure 5). When each folder is opened, all the individual records representing the person are displayed. Included with each record is a hyperlink to display the original source document. Above each Meta-Person is a group of "radio buttons" used to perform secondary queries of the individual for a specific purpose.

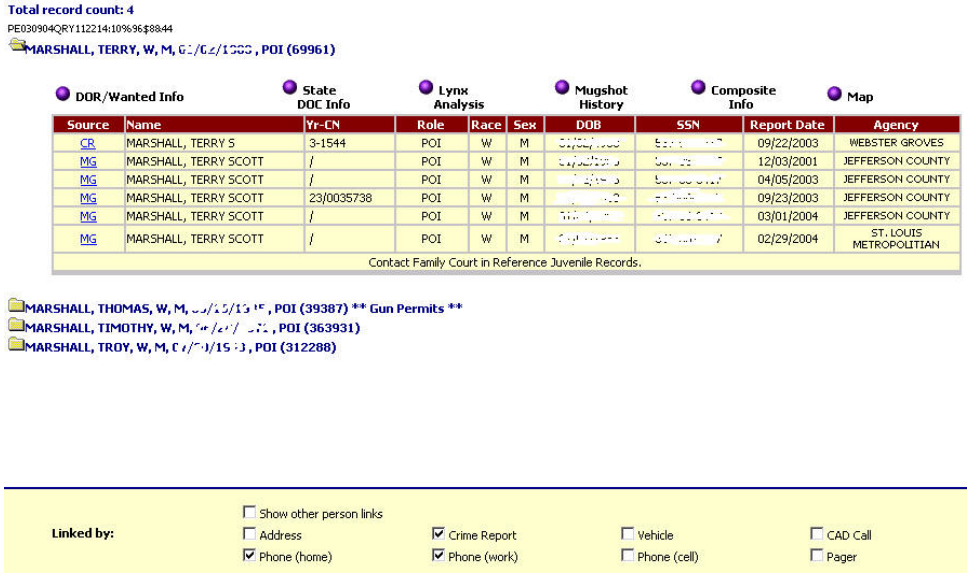

Figure 5. Document collection of Meta-Person

The "Composite Info" component summarizes the many attributes of the Meta-Person by providing the most recent image, address and arrest information. The CrimeMATRIX System through a scoring algorithm generates this consolidated view of an individual (Figure 6). Only the most recent data selected from multiple records is used to create this composite person. The webpage displays charge history, address history, deceased file, probation and parole status, gang membership, gun permit, and pawnshop activity information. The hyperlink for each of these links navigates the investigator back to any source document. Each pedigree data element folder displays a drop list and count of distinct value. For example, a "Person of Interest" (POI) may have seven distinct hair color values of brown, gray and blonde. The LYNX application performs a complete analysis of a person in seconds. In time studies conducted prior to the implementation of LYNX, a complete "person of interest" study would take 12.4 hours without providing any LYNX analysis information.

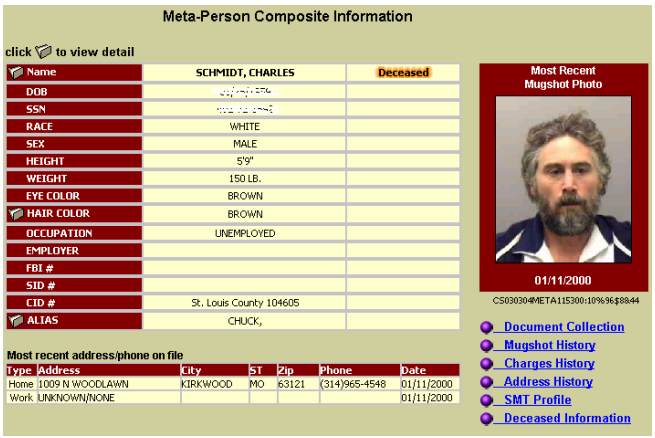

Figure 6. Meta-Person Composition Information

The LYNX application also integrates GIS functionality. The "Map" component is used to display the activity space of a Persons of Interest (Figure 7). An individual's home, work and prior arrest address information are plotted. Additional map layers displaying real estate data, municipal and beat boundaries, and aerial images are provided. The bottom frame of the webpage allows the investigator to display the source document that generated the map coordinates. This spatial representation of a person's activity space is an invaluable investigative tool when generating a potential pool of suspects who are active in a particular area.

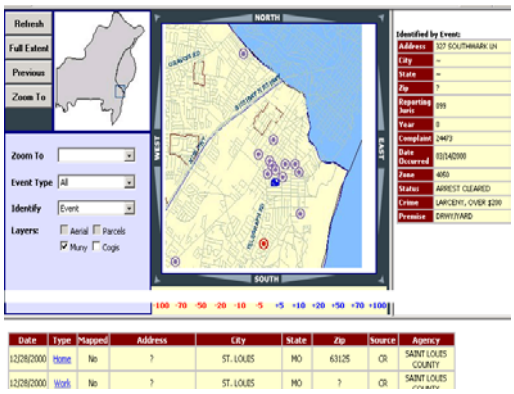

Figure 7. LYNX mapping component

Lynx also incorporates a mug shot imaging component into its search results (Figure 8). Individual mug shots, and associated booking data for each arrest are provided. A chronological mug shot history of the Meta-Person is provided to reveal changes in appearance. An integrated mug shot lineup feature is provided to create quick digital lineups. All lineup documents are saved for later retrieval by the investigator. All SMT (Scar, Marks and Tattoo) images are provided to the investigator and are searchable by image type and or content. Because suspects frequently change their appearance, these prior images with or without facial hair or differences in hairstyle are very useful for mug shot lineups.

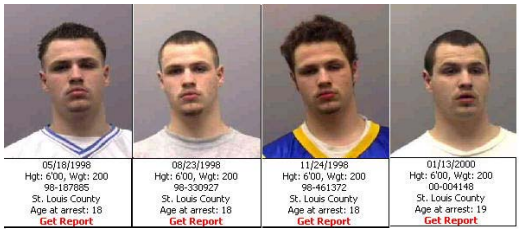

Figure 8. LYNX Mugshot history

The LYNX application interfaces to the MULES (Missouri Uniform Law Enforcement System) and NCIC (National Crime Information Center) for driver license and wanted information. This feature enables the investigator to perform real–time inquires of State and Federal data. The link provides the most current driver license image and driving status. It also queries the local, state and federal databases for wanted and probation information for the individual.

One of most interesting features of this application is the "LYNX Analysis" query. This component searches the CrimeMATRIX database and identifies any other individuals in the system that share a common crime, telephone number, vehicle or address. The component also displays victims and witnesses associated with a person of interest when linked through a crime. The query displays in a military matrix grid, each individual linked to the target "person of interest." In addition, the grid also displays all relationship between secondary individuals (Figure 9). For example, suspect A knows suspect B and C through a shared work address and suspect B and C share a common crime and telephone number.

By hovering the mouse over intersecting relationship cells, the application displays the two parties. Clicking on a cell will display the two parties and display how the individuals share an association. For example, David Fortune and Rockie Fortune share three associations, two for a set of common home numbers and one for sharing an arrest in a larceny offense. Table 4 summarize the application and time saving features for LYNX.

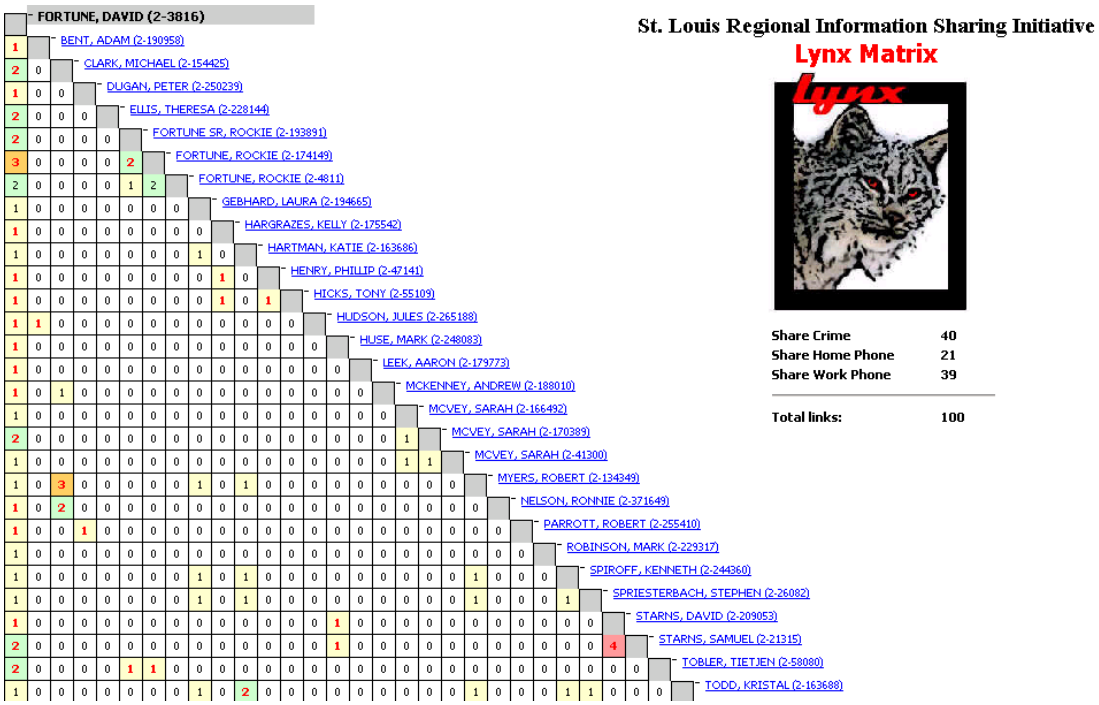

Figure 9. LYNX's association matrix

Table 4. LYNX Features

Application Features

- Browser-based application
- Integrates data from mug shot, sex offender, tickets, gang, police report, management, probation and parole, and gun permit systems
- Creates an association matrix of all individuals linked to the "person of interest" by telephone number, crime, employer, address, and vehicle.
- Provides a complete mug shot history
- Provides detailed descriptions and images of scars, marks, and tattoos
- Summarizes arrest and charge history
- Summarizes home, work and arrest location address history
- Summarizes employer and occupation history
- Redefines a person into a composite or meta-person with the latest pedigree, employment and address information
- Real-Time interfaces to State driver license images and wanted files. Also interfaces to NCIC wanted files
- Integrated in RAMS desktop application
- Integrated "WebMap" that displays a "person of interests" activity space (crime, home, and work locations)

Time Saving Features

- Reduces man hours while increasing efficiency
- Reduces time spent searching multiple information systems.

FUTURE DEVELOPMENTS

In addition to the RAMS and LYNX applications, a new set of browser-based mapping applications are under development. These applications will provide an easy means of displaying crime trends, recent probation & parole releases, sex offender work and home address information, and pawnshop monitoring data. The applications will operate on desktop computers and in our high-speed 3G (wireless network capable of transferring data at speeds of up to 384Kbps) mobile computing environment. The objective of these applications and future CrimeMATRIX system enhancements is to push out information to all levels of the law enforcement environment. Police officers on the street should have access to investigative information in their vehicles, which serve as their mobile office. Precinct/District Commanders, Police Chiefs and Police Planners should have access to investigative and crime trend data to establish proactive strategies to fighting crime.

These new mapping applications will utilize the ESRI ArcIMS development environment. They will be supported by an ESRI SDE (Spatial Data Engine) that is interfaced to the CrimeMATRIX database engine. We envision that this combination of a hybrid data collection and consolidation system with a "state of the art" mapping system allow us to take a more proactive approach to crime fighting models for the future.

REFERENCE

Environmental Systems Research Institute [ESRI]. *Map Object 2.0*, Redlands, CA., 1992.

Han, Jiawei and Kamber, Micheline, *Data Mining Concepts and Techniques*, San Francisco, CA: Morgan Kaufmann Publishers, 2001:13, 60.

Muhammad Qazzafi Balouch, *Information System and Database for Foreign Department Office,* Bachelor Thesis, Riga Technical University, Latvia:1999. http://www.geocities.com/rehanaq/bachelor.html

Power, D.J., "A Brief History of Decision Support Systems", posted at DSSResources.com, version 2.8, November 12, 2002, http://www.dssresources.com/history/dsshistory.html

Rossmo, Kim. *Geographic Profiling*, Boca Raton, FL: CRC Press LLC, 2000: 111.