THE DIGITAL CASE FILE: THE FUTURE OF FIGHTING CRIME WITH BIG DATA

Frank Hartle III, Robert Morris University, fchst270@mail.rmu.edu
Michael Parker, Robert Morris University, mapst149@mail.rmu.edu
Christopher Wydra, Robert Morris University, cawst157@mail.rmu.edu

ABSTRACT

Through history, American law enforcement has been a profession that necessitated the collection of a substantial and disparate amount of information. This information was and is stored in order to memorialize events, solve problems or crimes and identify, apprehend and prosecute lawbreakers. Until recently, leveraging this vast collection of data was a laborious and time consuming task. This article looks at current and emerging technologies and envisions a future where big data transforms the criminal justice system. The emergence of wearable technologies, new surveillance vehicles, and automated systems coupled with the power of large data warehousing creates an intriguing view into the future of crime fighting and prosecution. This content analysis evaluates the evolution of information management supporting law enforcement, culminating in a review of the literature supporting the digital case file. Further, this analysis purports research opportunities found within the content and literature to-date.

Keywords: Big Data, Artificial Intelligence (AI) and Information Technology (IT)

INTRODUCTION

Since the beginning of modern American policing, police officers, detectives or agents have endeavored to remove, prevent, or reduce crime and if unsuccessful to investigate the apprehend the offender. This necessitated the gathering of facts, identifiers and evidence. This information was and is stored in order to memorialize events, solve problems or crimes and identify, apprehend and prosecute lawbreakers. As more data is collected this information can be used for the identification of criminals or criminal groups, to identify patterns and anticipate crime and to identify the needed allocation of resources combat the crime. Manning (1992) scientifically described this gathering of information by stating:

“The police gather primary information or "raw data" that is then processed within policing for crime solving or closing the events to become secondary information. When processed twice, gathered, and formatted, it can move up the organization to become tertiary or "managerial" information. These forms of information and intelligence (information gathered for anticipated events, rather than gathered in response to an ongoing event) are realized and interact with police operational strategies (the allocation of resources to obtain a preventive, prospective, or reactive end). [15]

Until recently, leveraging this vast collection of data was a laborious and time consuming task. This paper seeks to highlight the current, multiple technology components that already exist in order to implement a local and state level data driven, digital criminal justice information systems and to further identify information management considerations in development. Further, this paper forecasts how current technology and big data could transform policing and the entirety of the criminal justice system. Finally, this paper underscores considerations in the development of a data driven, digital criminal justice information system, focusing on the key pillars of data management: collection, storage, exposure and extraction.

Digital Evolution of Policing

The adaptation of digital technology in policing has been forced by necessity and oversight [29, 6] President Lyndon Johnson’s 1967 Crime Commission report brought a new infusion of science and technology into policing which had stagnated prior to the great depression. Although military technology had exploded during this time, policing was not keeping pace. The Crime Commission report highlighted the need for the federal government to help local law enforcement agencies incorporate this new science and technology into law enforcement. The Omnibus Crime Control and Safe Streets Act of 1968 formed the Law Enforcement Assistance Administration (LEAA). This new administration began to fund computer driven command and control systems and urged centralized communications
modeled from the military [16,29]. After the recognition in the 1960’s that local law enforcement was falling behind technologically, several more events would force the profession into a greater reliance on computing and military technology; the war on crime, the war on drugs and the war on terror [6].

As crime rose drastically in the 1970’s and 80’s it became apparent that the amount of crime data being collected was overwhelming the ability of departments to collect, store, recall and discern important and related information for investigation and prosecution. The development of cheaper personal computers starting in the 1980’s along with simpler and more useful software programs enabled computer use to expand into two-thirds of police departments by the early 90’s [29]. However, having the ability to store and recall relevant information digitally was only part of the solution. O’Shea & Muscarello (1998) observed:

“Police detectives in large urban police departments are not likely to proactively discriminate crime patterns due to human information processing constraints and large crime data sets. The volume of criminal cases reported and the number of characteristics of each criminal incident make it virtually impossible to match and compare the extensive data involved over time. In order to overcome information processing deficiencies and identify at least some portion of pattern criminal incidents, police detectives construct various heuristics or decision shortcuts” (p.1) [18].

and

“At present, there are no automated methods of identifying patterned incidents. Each detective must manually scan hard copies of case reports on a board containing hundreds of reports of incidents. The detective must somehow keep stored in memory the thirty or so characteristics of each case report. He/she must then compare all characteristics of all cases against all other cases, looking for combinations of characteristics that are sufficiently similar to warrant classifying those reports a pattern. In Chicago, over a period of one year, approximately thirty thousand robberies are reported. Detectives seeking patterns in this sea of data are faced with a formidable task” (p.21) [18].

while earlier Manning (1992) stated:

Information in police departments can be best characterized as systematically decentralized. Often, primary data known to one officer are not available to other officers because of the personalized practices of data “storage.” Most of the information that exists in policing is primary data possessed by aggregated records or files or the information stored mentally by an officer. Some information is located in officers' case files, private notes, or log books. Only under certain specified conditions does it become universally understood, generally shared, reproducible knowledge held collectively by the records or institutional memory of the organization. The fragmented character of the knowledge possessed by officers is "designed" in part to protect officers from close day-to-day supervision, review, and discipline".(p.370)[15]

The proliferation of the internet and digital computing together with the increased federal funding towards combating crime and terrorism along with technology creep from the military allowed for more police centric data systems to be developed and instituted. This combined civilian and military technology has infiltrated its way into law enforcement and is beginning to change how police business is done. [6, 26, 9]

Today (2014), digital technology has impacted many aspects of life, including personal, social, and work environments. The “digital age” has rapidly grown and in recent years, there has been a rapid increase in the use of new technologies such as web-enabled portable devices, mobile technologies, and social media [1]. The advancement of technology has also affected how law enforcement officials combat and investigate crime. These current technologies can be very beneficial to law enforcement and can streamline the investigative process [9]

As stated, in the 1980s and 1990s, law enforcement had access to very simplistic technologies such as land line based telecommunications, and mobile technologies such as phone pagers and two-way radio communications. The technologies during those time periods were not very effective or useful to law enforcement. The “digital age” has transformed how crimes are committed and how they occur. The advancement of technologies has assisted law enforcement agencies in how they patrol, how they investigate criminal cases, and how they store and use evidence.
It may appear that police are at the right place at the right time, but in reality, computer programming and software advances pinpoint locations and times that a crime may occur [4].

**Policing and Evidentiary Procedures with the Advancement of Technologies**

Law enforcement must recognize that technologies must be utilized in order to police effectively in the “digital age”. The “digital age” and the advancement of technologies has broaden the scope of criminality and have provided more opportunity for people to commit crimes.

“When the new computer technologies were introduced, people began using computers for illegal purposes almost immediately. Previously, the people doing this had a wealth of knowledge and experience in high technology, but it is now not uncommon for computer infringement to be done by ordinary citizens who have only basic computer skills” (p. 141) [15].

Technology has changed the circumstances surrounding criminal activity. Computers, mobile technologies, and knowledge of the use of technology can be used to combat and investigate cyber-crimes.

The “digital age” and technology have also impacted law enforcement’s evidentiary procedures of collecting evidence. Important evidence may be stored on web based, computer based and mobile technologies. Due to these technologies and their capabilities of storing information, legal guidelines have changed on how law enforcement agents are legally able to retrieve and store evidence that is stored on technologies. Computer and mobile technology use has increased dramatically and so has criminal involvement [22]. Law enforcement must abide by the proper legal procedures guiding the retrieval and storage of digital evidence in order to effectively prosecute a criminal case. A legal structure is required that will support early detection and successful prosecutions. The legal structure and legal methods of legally collecting and storing digital evidence that is needed to properly investigate and prosecute have lagged behind technological and social changes [22].

**Criminal Investigation, Arrest, and Court Proceedings**

Criminal investigation, arrest, and court proceedings are known as the criminal justice process, which can be lengthy and time consuming with many integral parts. The criminal justice process has become too long, too expensive, and unpredictable, and is a part of a consequence of an intricate criminal procedure process imposed in pursuit of perfect justice [4]. Law enforcement agencies have adopted technologies to assist and aid in criminal investigations, arrests, and court proceedings, but in limited capacities. Even with the use of certain technologies, the criminal justice process remains a lengthy process. New technologies can be utilized by law enforcement agencies and prosecutors to expedite and solidify criminal cases.

Advanced technologies, such as software, computer models, and data base usage, have already been developed and can be used for the future of policing. Better technology, educated and trained law enforcement officials, and better community ties are significant factors for the future of policing. New and advanced technologies can transform policing from reactive to proactive and can streamline the criminal justice process by using technologies that already exist. Software and data-based policing have already been developed to calculate the probability of specific crimes occurring in specific locations. Data-based policing is the concept behind predictive policing and is relatively straightforward; the data particularly focuses on the time, distribution, and geography of past events. The data is stored in data-bases and then ferrets out patterns that would not be apparent to the human eye [5, 24, 27]. Data-based policing can be advantageous to the storage and retrieval of digital evidence, predictive policing, and solidifying criminal prosecution.

**Criminal Justice Transformation**

In 2010, a group of over 50 local, state, federal and international agencies came together to publish list of long-term research needs. These agencies included most federal agencies including all of the intelligence agencies and military services as well as many local, state and federal law enforcement agencies. The Interagency Council for Applied Homeland Security Technology (ICAHST) recommended the 20 areas of research needs in order to identify,
investigate and assist in the implementation of new science and technology to improve the nation’s counterterrorism and homeland security capabilities [7] Figure 1.

**Figure 1.** The Interagency Council for Applied Homeland Security Technology Long-Term Research Needs

Although it was only four years ago, much of the research requested by ICAHST is being done with many technologies making their way into commercial, military and governmental use. For example, technologies like Palatir software suites and Thermopile’s iHarvest, among others, have been developed to fit these needs. [20, 25]. Palatir’s software gives law enforcement agencies the tools to explore interconnected data and presents the information visually on graphical maps. The software was developed as an artificial intelligence tool that augments human thinking and enables organizations to make sense of massive amounts of disparate data. [22]. Palatir claims that it can search and access all law enforcement databases in one place, manage criminal cases in a single unified intelligence and investigative platform and investigate targets in the field and raise situational awareness to the command center [Palantir.com]. In essence it tracks how people think and presents the data automatically as visual data in real time. In addition, iHarvest claims to organize, collect and report big sets of proprietary criminal data and makes recommendations of related data. Just using these two software companies as examples, most of the ICAHST research needs can be checked from the list. This data is culled from many databases including open source online information, publically available paid services, private corporations and private government data collections [9].

Another example of an agile approach to meeting the criminal justice needs is the development of technology to replace the paper-based system within the court proceedings. One example of the technology, known as JusticeTech, will create a paper-on demand model to eliminate paperwork and streamline the process. JusticeTech allows the entire criminal justice profession to share a common digital case file. The software affords automated workflow between courts, prosecutors, law enforcement, and other related agencies [8].

In addition to the commercially available software systems there are many national databases that are leveraged in law enforcement. The National Criminal Information Center, the Criminal Justice Information Services and the National Law Enforcement Telecommunications System are a few of the major and more robust legacy systems in use and run by the Federal Bureau of Investigation. In addition to these each city, county and state have separate criminal and intelligence systems [FBI.GOV].
To alleviate multiple and duplicative systems the Federal Bureau of Investigation implemented the Data Integration Visualization System (DIVS) during calendar year 2010. The DIVS capability is a window to the future of federal criminal justice information management, which reveals a means of implementing nation-wide information systems supporting local, state and federal criminal justice. According to the federal government IT dashboard:

“The purpose of the DIVS investment is to make all FBI intelligence and investigative data available and searchable by its main stakeholders, the FBI agents and intelligence analysts, through a single, secure, web-based search and analysis capability. The goal of DIVS is to introduce and maintain the advanced technology capabilities that are required to store, present and analyze the electronic data collected by or provided to the FBI". [IT dashboard.gov].

The military is also updating its criminal justice systems. For example, the Department of the Navy has initiated a program to implement an end-to-end case management system governing the criminal justice process. According to the Naval Justice Information Systems Functional Requirements Document, this new capability will implement enterprise processes and software to support the end-to-end military criminal justice case management solution supporting the United States Navy (USN) and United States Marine Corps (USMC) Law Enforcement, Investigations, Judicial Actions, Command Actions, and Corrections communities. While the systems mentioned above will use many governmental databases, the collection and use of law enforcement only data has shifted from proprietary, government run data bases to a more robust blended approach utilizing both open source and private data. These include foreclosure information, twitter, Facebook and other social media, automatic license plate readers, communications companies, GPS, facial recognition software for commercial use, banking information and purchasing data to name a few [9].

The addition of new predictive policing systems coupled with mass surveillance systems and DNA databanks will add to the amount of data collected and give pause to those concerned with privacy. [9]. Police have historically used the military as a testing ground for police technologies [29]. The Military Cooperation with Law Enforcement Act (MCLEA) passed by Congress in 1981 and Program 1033 passed in 1997 enabled the transfer of military equipment and technology to law enforcement [6] Civilized military hardware has seeped into policing in the form of un-piloted drones, armored vehicles, weapons, riot gear, biometrics, augmented reality, wearable physiological indicators and surveillance equipment. This military technology is adding to the ways police are gathering data and intelligence.

The research areas as identified in the aforementioned analysis is transforming law enforcement through the utilization of information technology, civilized and military hardware and automated, big data software applications. These technologies enable information collection, processing, analysis and dissemination without much and sometimes no human interaction. Information Technology is the critical enabler in transforming the end-to-end criminal justice system from a primarily paper-based process to a digital, IT-enabled, automated one [7]. Waller and Fawcett state in the Journal of Business Logistics 2013 [29], “Big data” is the buzzword of the day. However, big data carries with it the opportunity to change business model design and day-to-day decision making that accompany emerging data analysis. Big Data is unique because of the volume, variety, and velocity of the data, which today is widely available and much less expensive to access and store. Volume can occur in many ways. There are more data because, among other reasons, the data are captured in more detail”. For

**Information Process (Collection, Processing, Analysis and Dissemination)**

These systems will require a vast amount of storage. Currently law enforcement collect crime reports, incident reports, contact reports, every 911 call, police radio and silent dispatch communication, citizen complaints, missing persons, evidence collection, crime lab results, digital evidence from computer crimes, fingerprints, photos, accident reports, tickets, electronic communication, arrest reports, vehicle registrations, and warrant information. This list is not exhaustive as police collect much more data than we think. It is hard to imagine how much more data a police department could collect however, leveraging the systems outlined above we can add video from wearable and dashboard cameras, biometrics, face recognition, police worn physiological monitors, DNA, automatic license plate reader cameras, patrol GPS, predictive analytics, crime mapping, connected public and private video feeds, social media, open source internet monitoring, drone video, and private commercial data. Now couple this data with the
federal databases and deploy the software applications that will organize, manipulate, present and predict in real time.

The information process is comprised of collection, processing, analysis, and dissemination [7]. As pointed out above, within each component of the information process, there is emerging IT capabilities that improve the criminal justice process. Information collection requires some extensive field level and headquarters data input for proprietary information but as discussed, connecting many data sources outside of the agency expands the search and intelligence ability without having to input, store or manipulate the data. At least not until it becomes evidence. Information collection is critical at the lowest echelon and contains both structured, unstructured semi-structured formats. Information collection technologies are the linchpin to success in capturing images, words, numbers, and documents at the point of event in real time. Database integration with real-time indexing into a single software system that mines data, connects information and makes links and predictions is the future of big data criminal justice. Using Palantir’s application platform as an example (Figure 2.) illustrates the potential for big data integration across the entire criminal justice system.

Figure 2. Source: www.Palantir.com

One aspect of employing a big data criminal justice system is the collection and storage of evidence. Unlike the corporate setting, law enforcement investigators are required to collect digital evidence unmolested [12,11]. Currently, police digital forensic examiners make copies or seize the system, computer, file or logs and store them as evidence. A second copy or mirror image is made to search and to manipulate for evidence [11]. In this paper’s future system, digital evidence would be created and stored in real-time by the agencies own system. Not only would evidence be created as it is collected it would be coming from a myriad of different systems creating the digital case file, as the crime is occurring. When system and programs are queried or automated systems make predictions they too must be stored as they become probable cause or at least reasonable suspicion and predicate any future investigation. As such these records are discoverable [11]. This database warehouse must be separate, and protected from the other systems. In addition, it must have the ability to index all the disparate evidentiary data to the appropriate case file, officer, prosecutor, victim, defendant, and more.

Bill Inmon and Richard Kimball are the authorities on data warehousing development, implementation and sustainment; however, each philosophy is drastically different from the other. The Inmon approach to data warehouse development has a number of key success factors for systems with structured and unstructured data, however, either approach can work depending on the project variance.
According to Inmon, a data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision-making process. Whereas, Kimball defines a data warehouse as a data warehouse is a copy of transaction data specifically structured for query and analysis. “Traditional data warehouse methodologies such as Kimball et al. (1998) follow a waterfall approach to data warehouse development with a relative long period before any product is delivered to the customer.” [3], p. 99. Goede explores the ability to leverage agile systems development methodologies in data warehouse implementation. In his research, Goede found that “agile system development methodologies are indeed suitable for the development of data warehouses”. [3], p. 105. Data warehousing is the cornerstone of the criminal justice information system development, but an agile approach to development should be considered. Given the relative instability of requirements analysis in criminal justice information capabilities and the need to deliver functionality in a phased approach, an agile methodology is preferred. However, for evidentiary purposes Kimball’s approach may be best. [3]

Law Enforcement data warehousing requires immediate storage and retrieval of data and information through the Extraction, Transformation, and Load (ETL) process and exposure of data through service oriented architecture (SOA). “With so much data surging through organizations, data managers need to be able to make fresh and relevant data as easily accessible in near-real time as possible, while stale data is moved out of the way to less expensive locations.” [17] p.14

Information dissemination is attained through the use of Data marts [7]. Data marts support data consumer visibility and decision-making and expose authoritative data and information from the warehouse. Data marts are consumer specific, real time data exposure tools and drive business analytics to decision makers [3].

Challenges

Technology is changing the way law enforcement operates, how grants requests are formatted, and what is requested in the local operating budget. Technologies funded today were not even common knowledge just a few years ago. Law enforcement officials need to stay current with ongoing technological developments. Today’s (2014) law enforcement officials need to be cognizant of developing technology but also to have a working knowledge of what technology can do for their agencies. Law enforcement must be skilled in acquiring technology through a variety of funding sources. Due to such technologies changing the way police operate, law enforcement officials are changing their purchasing priorities as well, dedicating funds either from grants or from their operating budgets to keep their agencies technologically up to date [20].

Many law enforcement agencies have developed automated and system generated capabilities supporting criminal justice functions like the software and hardware technologies mentioned above. Soat discovered “employing off-the-shelf technology and developed by third-party service providers, many of these systems are standalone, which has led to another situation familiar to corporate IT – silos of data.” [28] p. 42. These silos of data reveal a challenge for criminal justice information management and require controlled integration and governance. There are a number of implementation challenges to consider when developing an end-to-end system solution in support of the criminal justice process. Requirements management is critical to understanding and controlling the consequences and prioritization of requirements. Change, especially technology change, requires direct management to support a successful transition. A number of cost drivers must be analyzed for evaluation of best business practices security and system/capability solutions. Internal controls over data and information processing is critical to the success of this transformational effort. Fourth Amendment Rights, Freedom of Information Act (FOIA) and Data Management Rights (DMR) are other critical challenges to collecting and storing big data from many sources while maintaining the civil rights of those whom law officer are sworn to protect.

Other challenges exist but are not insurmountable. These include, bandwidth for wearable and autonomous systems, recruiting and maintaining the technical skill set necessary to run the systems, and a robust wireless backbone to run a citywide system. Already systems like pCell are being developed to speed wireless up to 1000 times their current speed allowing large data streams to travel from these new technologies [19]. New programs were implemented that pay off school loans for highly technical people to work in governmental positions [26]. Much like the military pays for college so would this system that allows for some of the brightest to cut their teeth with a local government for a
few years. Finally, mesh networking are being developed to replace standard landline corporate and governmental IT backbones [27].

CONCLUSIONS

Villasenor, 2011 opined that in the near future, it will be possible and cost effective for the government to record everything anyone says or does. A scary thought considering the potential for misuse and abuse. However, in the right hands and under strict oversight, systems that were outlined above could be utilized to keep society safer from criminals and terrorist, provide a first person account of any situation including police use-of-force and to streamline the criminal justice system [10].

An intriguing scenario can be imagined when big data and the appropriate systems are in place. One could conceive a robbery taking place in an urban area. Predictive analysis has necessitated that more officers have been assigned to the area. As officer move towards the location a description is broadcast. At the same time the autonomous criminal justice systems begin their work. The systems automatically cull all license plate information from the Automatic License Plate Readers (ALPR) in the area within the last hour and search it against known robbery suspects, it also looks for similar plates that have been through the area in the last week and correlates the information to see if any matching suspects may have been casing the business. At the same time the automated system request and gather video from local private and public CCTV systems including the business that was robbed. Using this video, the system begins to run facial recognition programs for people in the area of the robbery at the time. A suspect is identified through facial recognition and correlated with the ALPR. Rooftop drones are launched and track the suspect as he runs from the scene. Police arrive on the scene but have been notified enroute, by the automated system, that the suspect is known to resist arrest. After a brief scuffle the suspect is arrested and a weapon is recovered.

At court the scene is very different than we are used to. Verbal reconstruction of the defense and prosecution is replaced by a time line video presentation of the crime with all the digital evidence presented beside the video. Witnesses are located using video and social media filters to identify first hand witnesses. Their posts are presented as evidence as they happen alongside the video. Police officer, witness and victim statements are played from the scene where they recorded. Physical evidence is presented but merely accents the real evidence as the jury watches the suspect commit the crime and follows as he is tracked and views his apprehension.

The suspect claims that excessive force was used the police. The DA and internal affairs use the vehicle, drone, body worn cameras video and physiological monitors unsubstantiated the claim.

The future of big data infused into the criminal justice system is exciting and a bit chilling. Appropriate oversight and fourth amendment protections must be part of any complete and functional system. This article looked at current and emerging technologies and envisions a future where big data transforms the criminal justice system. The emergence of wearable technologies, new surveillance vehicles, and automated systems coupled with the power of large data warehousing creates an interesting view into the future of crime fighting and prosecution.

This content analysis provides a rich collection of the evolution of information management supporting law enforcement over several decades and numerous information technology (IT) advancements. Through this content analysis, an opportunity to pursue research in the law enforcement utilizing cutting edge IT capabilities and data warehousing techniques is revealed. A specific research opportunity purported is a quantitative study of the impacts of the information process (collection, processing, analysis and dissemination) supporting law enforcement cases by comparing the utilization of big data strategies, data warehouse implementation and data mart utilization. Another research opportunity is a qualitative study – phenomenology – focused on the study of a phenomenological impact of implementing a big data strategy within a specific law enforcement agency.
REFERENCES


