Information Sharing and Collaboration Policies within Government Agencies

Homa Atabakhsh[1], Catherine Larson[1], Tim Petersen[2],
Chuck Violette[2], and Hsinchun Chen[1]

[1] Department of Management Information Systems
University of Arizona, Tucson, AZ 85721, USA
{homa,cal,hchen}@eller.arizona.edu
[2] Tucson Police Department
270 S. Stone Avenue, Tucson, AZ 85701, USA
Tim.Petersen@tucsonaz.gov, cviolet1@ci.tucson.az.us

Abstract. This paper describes the necessity for government agencies to share data as well as obstacles to overcome in order to achieve information sharing. We study two domains: law enforcement and disease informatics. Some of the ways in which we were able to overcome the obstacles, such as data security and privacy issues, are explained. We conclude by highlighting the lessons learned while working towards our goals.

1 Introduction

The need for information sharing between government agencies has been highlighted in the post-9/11 scrutiny of terrorism events. Historically information sharing between law enforcement agencies has occurred in a very limited manner: ordinarily, only by person to person or case by case basis. In our current age of high mobility and increasing availability of technology, criminals are able to take advantage of the fact that limited information sharing between law enforcement jurisdictions reduces the likelihood of getting caught.

In the public health realm, health-related data is also shared on a case by case basis, between doctors, hospitals, laboratories, insurance providers, and others involved in the provision of healthcare services. Public health data is often shared at an aggregated level. The mechanisms used for data transmission (such as fax, telephone, email and other methods) are often cumbersome and rudimentary, and do not support analysis, disease management or prediction at a regional or national level [13]. Data that is published on the Web is not necessarily interactive. Data sharing is guided and restricted by both a plethora of legal regulations as well as informal guidelines such as those in private practices. Aggregated data is reported to local and/or state and national agencies but not in real time.

There are many roadblocks to overcome in the process of sharing information between agencies, whether law enforcement or public health. Most of these roadblocks are no longer technical in nature. Political and social barriers are far and away the greater obstacles to overcome. Factors such as variations in state and federal laws, city codes and department policies affect this process greatly. We have experienced

H. Chen et al. (Eds.): ISI 2004, LNCS 3073, pp. 467–475, 2004.

that as the size of the involved bureaucracies grows, the size of the barriers increases proportionally.

In this paper, two case studies are described, one in law enforcement and one in disease information. For each case study, the need for information sharing between government agencies is described followed by obstacles encountered and some of the ways in which we have overcome these obstacles. Data privacy and security issues are also described. Technical solutions for solving security problems present a research topic of their own and are out of the scope of this paper. We conclude with some of the lessons learned during this process.

2 Background and Motivation: Information Sharing between Government Agencies

2.1 Case Study: Law Enforcement

2.1.1 Introduction. In response to the September 11 terrorist attacks, major government efforts to modernize federal law enforcement authorities' intelligence collection and processing capabilities have been initiated. At the state and local levels, crime and police report data are rapidly migrating from paper records to automated records management systems in recent years, making them increasingly accessible.

However, despite the increasing availability of data, many challenges continue to hinder effective use of law enforcement data and knowledge, in turn limiting crime-fighting capabilities of related government agencies. For instance, most local police have database systems used by their own personnel, but lack an efficient manner in which to share information with other agencies [8, 10, 12]. More importantly, the tools necessary to retrieve, filter, integrate, and intelligently present relevant information have not yet been sufficiently refined.

As part of nationwide, ongoing digital government initiatives, COPLINK [3, 4, 6] is an integrated information and knowledge management environment aimed at meeting some of these challenges. Funded by the National Institute of Justice and the National Science Foundation, a prototype for COPLINK was initially developed at the University of Arizona's Artificial Intelligence Lab in collaboration with the Tucson Police Department (TPD) and Phoenix Police Department (PPD). COPLINK was developed into a product by Knowledge Computing Corporation (KCC) and deployed in approximately one hundred law enforcement agencies nationwide [1, 11].

The main goal of COPLINK is to develop information and knowledge management systems technologies and methodology appropriate for capturing, accessing, analyzing, visualizing, and sharing law enforcement related information. The COPLINK project has already bridged gaps between law enforcement agencies by allowing secure access by officers of some of the participating agencies. COPLINK has already shown its capabilities in the area of data sharing. As an example, the following is a quote from one of the Tucson Police Department (TPD) officers using COPLINK.

“COPLINK saved me tens of hours and the possible closure of a child rape case. My case involved a suspect only known as (NAME WITHHELD). We did not know how to spell his name and no address. Using COPLINK 's wildcard search I located a lost wallet report to the suspect's father and at the list location I found the suspect's vehicle. Several months later San Diego has a DNA hit on a child kidnap/rape case and I was able to provide them with suspect information. If COPLINK had been connected to Phoenix [this investigation occurred prior to TPD-PPD COPLINK connection] I would have been able to pull his photograph and fingerprints from an arrest I learned about later. This person has a warrant for his arrest. Without COPLINK I probably would have closed the case as unsolved."

2.1.2 Data Sharing for Law Enforcement: From Knowledge Discovery and Dissemination to Border Safe. The Knowledge Discovery and Dissemination (KDD) project (funded by the National Science Foundation) and the BorderSafe project (funded by the Department of Homeland Security) are recent AI Lab initiatives that encompass collaborative efforts between the University of Arizona's AI Lab, law enforcement agencies in Arizona such as TPD, PPD, Pima County Sheriff's Department (PCSD) and Tucson Customs and Border Patrol (CBP) as well as San Diego ARJIS (Automated Regional Justice Systems) and San Diego Supercomputing Center (SDSC). These projects expand on existing partnerships and technologies in addition to breaking new ground in both areas.

Federal, State and local regulations require that agreements between agencies within their respective jurisdictions receive advanced approval from their governing hierarchy. This precludes informal information sharing agreements between those agencies. We found that requirements varied from agency to agency according to the statutes by which they were governed.

For instance, the ordinances governing information sharing by the city of Tucson varied somewhat from those governing the city of Phoenix. This necessitated numerous attempts and passes at proposed documents by each cities law enforcement and legal staff before a final draft could be settled upon for approval by the city councils. We found in general that similar language existed in the ordinances and statutes governing this exchange but the process varied significantly enough to require modification in almost every case. It appears as though the size of the jurisdiction is proportional to the level of bureaucracy required.

Our initial experience in developing an agreement between agencies in Arizona and agencies in California is following this premise. Negotiating a contract between University of Arizona and ARJIS (Automated Regional Justice Information System) of Southern California required two to three months of negotiation between legal staff, contract specialists and agency officials. We are hopeful that many of the solutions to barriers in that process may be applied to the formation of formal agreements for information sharing with other agencies that cross state boundaries.

In an effort to facilitate data sharing between the agencies involved in the KDD and Border Safe projects, TPD has recently written an Intergovernmental Agreement (IGA) that will be signed between different law enforcement agencies. This IGA was condensed from MOU's (memorandum of understanding), policies and agreements that previously existed in various forms between numerous agencies. The IGA was drafted in as generic a manner as possible, including language from those laws, but excluding reference to any particular chapter or section. That allowed the required verbiage to exist in the document without being specific to any jurisdiction.

Sharing of information between agencies with disparate information systems has also lead to bridging boundaries between software vendors and agencies (their customers). We took care not to violate licensing terms by insuring that non-disclosure agreements existed and that contract language assured compliance with the vendors' licensing policies.

There is now a working VPN between ARJIS and TPD over which a small group of investigators from each department have begun querying the others' dataset. One day after implementing this connection a TPD crime analyst investigating information about a suspect from California involved in a TPD incident was able to obtain mug photos and drivers license photos from the ARJIS connection. The information and photos will be used to create a bulletin warning of officers' safety issues related to this person. Also, the gang sergeant at TPD has already found information in the ARJIS dataset relevant to criminals he is currently investigating in Tucson.

The activities of many individual criminals clearly span across multiple jurisdictions. There is obvious value in the sharing of data between jurisdictional authorities. The Border Safe project has begun exploration of the integration of information across datasets provided by multiple jurisdictions in the southwest United States. Many local incidents involve vehicles registered in other jurisdictions. This analysis is even more powerful when border-crossing information can be merged with networks of incident information. The Tucson sector of the Bureau of Customs and Border Protection (CBP) has shared a list of border crosser license plates with the AI lab for analytical research. Tying together criminals, their vehicles and activity histories can produce powerful tools to create leads for criminal investigation.

2.1.3 Access and Control Issues for Law Enforcement Data. The information being shared in COPLINK, KDD, and the BorderSafe projects is for law enforcement use only and is compiled from documented law enforcement contacts. Data sources such as medical records, bank records, credit histories, and other non-law enforcement related information are not (and will not be) a part of these projects. The sharing of law enforcement information is vital in solving crime and not intended to provide broad access to private information.

In any data sharing initiative, it is essential to make sure that the data shared between agencies is secure and that the privacy of individuals is respected. We have taken the necessary measures to ensure data privacy and security. It is important to note that the data shared between agencies through the initiatives discussed above, contains only law enforcement data and is available only to individuals screened by these agencies using TPD Background Check, Employee Non-Disclosure Agreement (NDA) and the TOC (terminal operator certificate) test.

Currently all personnel who have access to law enforcement data fill out background forms provided by TPD and have their fingerprints taken at TPD. They also sign a non-disclosure agreement provided by TPD. In addition, they take a TOC (terminal operator certificate) test every year. The background information and fingerprints are then checked by TPD investigators to ensure the lack of involvement in criminal activity and for verification of identity.

In addition to the above forms and test, all law enforcement data in the University of Arizona's AI Lab reside behind a software firewall and in a secure room accessible only by activated cards to those who have met the above criteria (i.e., background check, NDA and TOC test approved by TPD). As soon as an employee stops working on projects related to law enforcement data, their card is de-activated. However, the NDA is perpetual and remains in effect even after an employee leaves.

2.2 Case Study: Developing a National Infectious Disease Information Infrastructure (NDII): An Experiment in West Nile Virus and Botulism (WNV-Bot)

2.2.1 Introduction. Public health agencies have long recognized the need to develop new models for partnerships and improve data and information sharing about disease outbreaks, particularly since the anthrax bioterrorism event of 2001 [2] and the recent SARS outbreak [5]. The increasing awareness of bioterrorism as a threat to the health and safety of U.S. citizens has intensified recognition of the need to develop disease surveillance and data sharing capabilities that support real time analysis, and facilitate communications about outbreaks, whether naturally occurred or human-caused [7]. As public health records migrate from paper-based to automated records management systems, it becomes even more critical to design system architectures that are flexible, scalable and inter-operative.

As part of broader national efforts to improve information sharing about disease outbreaks, the National Science Foundation awarded funding to the University of Arizona's Artificial Intelligence Lab in collaboration with its partners, the New York State Department of Health (NYDOH), and the California Department of Health Services (CADHS). The National Biological Information Infrastructure/National Wildlife Health Center of the United States Geological Survey is also participating as a data provider. The partnership, under the direction of the Disease Informatics Senior Coordination Committee, has developed the West Nile Virus-Botulism (WNV-Bot) portal as a prototype of a national infectious disease information infrastructure.

The prototype is intended to apply advanced informatics techniques and demonstrate that an architecture for capturing, accessing, analyzing, and visualizing disease-related data from multiple sources can be successfully created and can be extended to include real-time reporting and alerting capabilities and the ability to interoperate with other systems. The research objectives of the portal include:

- Demonstrating and assessing the technical feasibility and scalability of an infectious disease information sharing, alerting, and analysis framework
- Developing and assessing advanced data mining and visualization techniques for infectious disease data analysis and predictive modeling
- Identifying important technical and policy-related challenges to developing a national infectious disease information infrastructure

Test data belonging to participating state agencies has been integrated into the portal and includes scrubbed and abbreviated case records for humans and birds, mosquito counts, chicken sera, dead bird data, and botulism data. Additional data has been integrated into the portal to allow spatial and temporal visualization and analysis, and to support experimentation with hot-spot analysis. Such data includes, for example, vegetation, temperature, rainfall, bird migration, and unemployment. The inclusion of these datasets supports the analysis of disease outbreaks against the backdrop of their environmental or demographic factors.

When fully developed, the WNV-Bot portal prototype will be enhanced to include other diseases, and will be made available to public health officials for extensive testing. If successful, it will be expanded to the national level with a different funding and governance structure.

2.2.2 Access and Control Issues for Disease-Related Data. In the U.S., the states exercise the primary power for protecting citizen health, but the federal government plays an important role in gathering and disseminating public health data through a plethora of agencies; and city and county level health agencies also fulfill information gathering and other functions [9]. As with law enforcement data, state and local regulations require that agreements between agencies receive advanced approval from their governing hierarchy, precluding informal information sharing agreements between those agencies. We also found that requirements varied from agency to agency according to the statutes, regulations or policies by which they are governed. For example, CADHS, NYDOH and the University of Arizona each have different regulations regarding the treatment of confidentiality and for how long data may be kept.

Similar to the law enforcement case study, the WNV-Bot portal project is now attempting to make its agreement structure as flexible and scalable as its system architecture. A generic memorandum of understanding (MOU) has been collaboratively drafted and principles agreed to that meet the requirements of current partners. The MOU specifies that data must be returned or destroyed after five years; that the data may not be shared with anyone outside of the agreement; that collaborating partners retain data ownership, but mutually own data analysis; and that the agreement may be expanded to include diseases other than those currently specified.

In drafting this more generic MOU, we hope to meet the requirements of future collaborators as well as current partners. As a new agency, organization, or other entity joins the partnership, it will be required to agree with the terms and conditions specified in the MOU. Confidential disclosure agreements specifying the level of access privilege will be signed by individuals as their organizations join the project.

Challenges to drafting an MOU that all parties can agree on has included meeting differing state requirements, the ineffectiveness of communication methods used to resolve differences and negotiate legal compromises, and heavy workloads and shifting priorities within the various agencies and offices. Differing state requirements can sometimes be negotiated and compromises made, but emailing and faxing are not necessarily effective communication means for resolving sensitive issues. Instead, having the contracting officers from the participating offices actually talk on the telephone point by point has been more effective and we believe, more apt to result in a workable agreement. Heavy workloads and shifting priorities may not be in the control of the project team, but daily attention through phone calls has proven to be the best method for ensuring that the paperwork does not get continually placed at the bottom of the priority list.

Resolving these access and control issues through a small prototype is, of course, far more workable than trying to resolve them directly at a national level.

Disease outbreaks do not recognize political boundaries, and may not necessarily be confined by geographic boundaries. *Overt* (naturally occurring) outbreaks may be initially noticed and managed by a public health agency. *Covert* (human-caused) outbreaks may not be initially recognized as such and thus may first be managed by a public health agency followed by a law enforcement agency [2]. Overt diseases can often be tracked through a geographic spread, enabled by animal or human movement. The possibility of a covert disease outbreak triggered by large-scale bioterrorism events that are not geographically co-located makes sharing at the national and even global levels even more critical, and the negotiation of data-sharing agreements even more complex. The WNV-Bot portal project will serve as a test case for identifying legal and contractual issues around the sharing of sensitive information.

2.2.3 Privacy Issues for Disease-related Data. As of this writing, formal agreements that will allow the sharing of real-time data are in the process of being signed. Data currently in use in the portal is test data. The information in the WNV-Bot portal is for use by participating public health officials and research partners only. As with the law enforcement data, it is essential to protect the privacy of individuals whose data may be included in the portal. The portal is to be used only for disease tracking and analysis, across jurisdictions, and is not intended to provide broad access to private information.

The design for user data access control is modeled after the New York State Health Information Network/Health Provider Network (HIN/HPN). Data providers and users (including their institutions) need to register with the Portal, and the data providers assign access privilege level to individual users. Predefined access privilege levels include:

- Aggregated data (for example, access to weekly data at county level only)
- Detailed data (access may still be restricted to certain fields)

Access and authorization is being managed through a Java-based User Access Control API and access privilege definition. Data is loaded into the portal using Secure Sockets Layer encryption. Once formal agreements are reached, real data can be loaded into the portal for extensive user testing by project partners.

It is envisioned that, once fully implemented, a national disease information sharing infrastructure would be used by public health professionals, researchers, policy makers, law enforcement agencies and other users with a need for the information.

3 Lessons Learned and Conclusion

Our planned time for reaching agreements on contracts, MOUs, CDAs and IGAs and other required legal documentation was originally far too optimistic in both our law enforcement and disease informatics case studies. We found that agreement processes often took in excess of six months to complete. The upside has been the creation of a framework and templates, which will hopefully speed and smooth the process in future endeavors. We anticipate providing free access to models of the types of documentation we used and developed, which may facilitate duplication of this process by others. It is very important to keep security of information among the highest priorities to ensure confidence and trust between participants.

Throughout this process we have found it necessary and helpful to have a single person within each agency responsible for ensuring that the process does not slow down or stop. We needed daily emails, phone calls and face-to-face contact in order to follow up on commitments, check on progress and move the project forward.

Our initial research and experience has bolstered our belief that as our society becomes more mobile and criminal activity follows that trend, the need for law enforcement and public health officials to keep pace is reinforced. Law enforcement must be able to follow the activities of criminals beyond the boundaries of a city, county or state. New criminal venues such as cyber-crime, identity theft and international and domestic terrorism require that the law enforcement community respond with measures such as better information access among officers. Increased mobility also means that public health officials must be able to track the spread of microbes around the world. The hazards posed by new, naturally occurring diseases as well as the threat of bioterrorism events requires that public health agencies actively develop the means to share information and analysis more quickly and accurately than ever before. Given the possibility of covert outbreaks, partnerships between law enforcement and public health, and the ability to share information across domains and not just across jurisdictions, becomes even more critical.

Acknowledgements

These projects have primarily been funded by the following grants:

- NSF, Digital Government Program, "COPLINK Center: Information and Knowledge Management for Law Enforcement," #9983304, July, 2000-June, 2003.
- NSF, Knowledge Discovery and Dissemination (KDD) # 9983304, June 2003-March 2004 and October 2003 – March 2004.
- NSF, ITR: "COPLINK Center for Intelligence and Security Informatics Research – "A Crime Data Mining Approach to Developing Border Safe Research." Sept. 1, 2003 – Aug. 31, 2004.
- DHS / CNRI: "Border Safe," Sept. 2003 – Nov. 2004.
- National Institute of Justice, "COPLINK: Database Integration and Access for a Law Enforcement Intranet," July 1997-January 2000.

We would also like to thank: Members at the University of Arizona Artificial Intelligence Lab and especially the COPLINK and West Nile Virus-Botulism portal teams, Tucson Police Department, Phoenix Police Department, Pima County Sheriff's Department, Tucson Customs and Border Protection, ARJIS (Automated Regional Justice Information System), SDSC (San Diego Supercomputing Center), and our partners at New York State Department of Health, California Department of Health Services, and the National Wildlife Health Center.

References

1. *Anchorage Daily News*, November 23, 2003, "Software Joins Cops on the Beat," COPLINK program links databases, speeds police investigations in the state of Alaska.
2. Butler, J., L. Cohen Mitchell, C. R. Friedman, R. M. Scripp, C. G. Watz. "Collaboration between Public Health and Law Enforcement: New Paradigms and Partnerships for Bioterrorism Planning and Response. " *Emerging Infectious Diseases* 8(100): 1152-1156.
3. Chen H., Jenny Schroeder, R. V. Hauck, L. Ridgeway, H. Atabakhsh, H. Gupta, C. Boarman, K. Rasmussen, A. W. Clements (2002), "COPLINK Connect: Information and Knowledge Management for Law Enforcement"; *Decision Support Systems*, 34: pp. 271-285.
4. Chen H., Zeng D., Atabakhsh H., Wyzga W., Schroeder J. "COPLINK: Managing Law Enforcement Data and Knowledge." *Communication of the ACM. Special Issue on Digital Government: Technologies and Practices*. 2003

5. Dignan, L. (2003), "Diagnosis: Disconnected." *Baseline.* http://www.baselinemag.com/print_article/0, 3668, a=41305, 00.asp, accessed Feb. 8, 2004.
6. Hauck R., H. Atabakhsh, P. Ongvasith, H. Gupta, and H. Chen (2002), "Using COPLINK to Analyze Criminal Justice Data," *IEEE Computer*. March 2002.
7. Henderson, D. A. (1999), "The Looming Threat of Bioterrorism, " *Science* (283), Feb. 26: 1279-1282.
8. Hoogeveen, M. J. & K. van der Meer (1994). "Integration of Information Retrieval and Database Management in Support of Multimedia Police Work, " *Journal of Information Science* 20(2): 79-87.
9. Institute of Medicine (1988). *The Future of Public Health.* http://www.nap.edu/books/0309038308/html/, accessed February 8, 2004.
10. Lingerfelt, J. (1997). "Technology As A force Multiplier, " *Proceedings of the Conference in Technology Community Policing*. National Law Enforcement and Corrections Technology Center.
11. Los Angeles Daily News, December 6, 2003, "Cops Could Hit the Links Soon: New Search Engine Would Catalog, Interpret Data for Investigations,"
12. Pliant, L. (1996). "High-technology Solutions, " *The Police Chief* 5(38): 38-51.
13. Thacker, S.B., K. Choi, P.S. Brachman (1983). "The Surveillance of Infectious Diseases, " *JAMA* 249(9): 81-5.