Managing digital forensic knowledge an applied approach

David P. Biros  
*Oklahoma State University*

Mark Weiser  
*Oklahoma State University*

John Witfield  
*Air Force Institute of Technology*

Follow this and additional works at: https://ro.ecu.edu.au/ADF

Part of the Computer Sciences Commons

**Recommended Citation**


DOI: 10.4225/75/57ad592f7ff2f

Managing Digital Forensic Knowledge
An Applied Approach

David P. Biros and Mark Weiser, Oklahoma State University and Edith Cowan University
david.biros@okstate.edu, weiser@okstate.edu
John Whitfield, Air Force Institute of Technology
John.Whitfield@afit.edu

Abstract

The science of digital forensics is continually changing as technological advances are made and new digital devices are developed. This environment forces analysts to regularly extend their skills with training and frequent research to develop new and admissible techniques. Unfortunately, the same and similar methods are re-discovered by other analysts who are unaware of earlier peer efforts. The situation is aggravated by a nearly universal backlog in qualified digital forensics facilities. This leaves little time for communication between analysts even within a single agency.

To address these issues and facilitate an increase in efficiency across all law enforcement agencies, we apply the lessons of knowledge management to digital forensics and extend them with special characteristics required by the law enforcement profession. The result is the development of the National Repository of Digital Forensic Intelligence. This system has been implemented in the largest accredited digital forensics lab in the world and is currently being extended to many other local, state, and federal agencies to increase effectiveness and efficiency among analysts.

INTRODUCTION

Rarely does a day pass that we are not made aware of a significant computer security breach that potentially puts our private information, finances, or even personal security at risk. The anonymity and freedom that the public demands in digital interactions is the same anonymity and freedom exploited by those who wish to do us harm. No well-connected computer is perfectly secure, so there is a balancing act between ease of use for legitimate transactions and security against illegitimate actions. Because technical counter-measures and training are insufficient to offset breaches, existing and new laws have been applied to the digital world to allow legal pursuit of those who seek to violate our digital worlds.

Digital devices may be the object of a crime, or an instrument to commit a criminal act. More often than not, however, digital evidence is being brought to bear in crimes that are not computer-based in any way. Even beat cops are being trained to ensure that digital evidence is preserved and seized in a manner acceptable to the courts. The increased awareness of the value of this evidence has resulted in a greater demand for forensic analysts and a need for them to work more efficiently and effectively. The bulk of leading-edge digital forensic knowledge is held in the minds of the analysts. The combination of a growing backlog and requirement for continuous innovation to keep up, however, has left little time for collaboration between examiners. Managing their knowledge in a way that limits redundant creation and allows sharing and efficient use among law enforcement agencies is the only way that these critical techniques can be properly leveraged.

This paper frames digital forensics as a knowledge management issue and applies some special characteristics of law enforcement. It then describes and updates progress on an ongoing collaboration between Oklahoma State University’s Center for Telecommunications and Network Security (CTANS) and the United States’ Defense Cyber Crime Center (DC3) to develop the National Repository of Digital Forensic Intelligence (NRDFI). This system, now in beta testing with multiple agencies, has the potential to rapidly and centrally make available
forensic discoveries throughout the DOD and law enforcement, without exposing those techniques to those who could exploit them for criminal activity. The goal is to provide a conduit for sensitive and relevant information interchange in a manner tailored to the needs of forensic analysts and law enforcement.

APPLICATION OF KNOWLEDGE MANAGEMENT

Digital Forensics is defined (Biros and Weiser, 2006) as “Scientific knowledge and methods applied to the identification, collection, preservation, examination, and analysis of information stored or transmitted in binary form in a manner acceptable for application in legal matters.” Meaning that all facets of identification, collection, preservation, examination, and analysis, must be verifiable and repeatable, and the results generally accepted by the digital forensic community. The rapidly changing nature of digital technology makes “general acceptance” difficult to attain. Reusing discoveries from other law enforcement agencies that have been successfully presented and accepted in a court is critical to gaining legal admissibility of the techniques.

Although law enforcement has some special characteristics, the sharing of knowledge between experts is important in many organizations. Knowledge and the ability to marshal and deploy knowledge across an organization are key factors for an organization’s competitive advantage (Vizcaíno, Soto, Portillo, & Piattini, 2007; Vouros, 2003; Teece, 1998; Tsai & Ghoshal, 1999). In order for organizations to remain competitive, knowledge management systems (KMSs) have been designed to manage an organization’s knowledge (Vizcaíno et al., 2007). In light of this, knowledge management systems are becoming ubiquitous in today’s corporations (Davenport & Prusak, 1998). KMSs are tools that affect the management of knowledge and are manifested in a variety of implementations including document repositories, expertise databases, discussion lists, and context-specific retrieval systems incorporating collaborative filtering technologies (Hahn & Subramani, 2000). The main objective of a KMS is to support the creation, transfer, and application of knowledge in organizations (Bock, Zmud, Kim, & Lee, 2005; Kalkhanhalli, Tan, & Wei, 2005). Alavi and Leidner (2001) defined a KMS as an information technology based system developed to support and enhance the processes of knowledge creation, storage/retrieval, transfer, and application.

KMS encompass a variety of technology-based initiatives such as the creation of databases of experts and expertise profiling and the hardwiring of social networks to aid access to resources of non-colocated individuals (Davenport & Prusak, 1998; Pickering & King, 1995). The primary focus of many of the KMS efforts has been on developing new applications of information technology such as data warehousing and document repositories linked to search engines to support the digital capture, storage, retrieval and distribution of an organization’s explicitly documented knowledge (Hahn & Subramani, 2000). Today’s KMSs store vast amounts of information and serve a variety of issues such as the creation and acquisition of knowledge in organizations, the storage and retrieval of available knowledge, and the sharing of knowledge among individuals and organizations, while they address the needs of an individual to interpret and reason about collective knowledge (Tiwana, 2000; Fahei, Srivastava, & Smith, 2001; Shin, Holden, & Schmidt, 2001).

KNOWLEDGE MANAGEMENT SYSTEM ACCEPTABILITY

Recent literature in the information systems field extols the virtue of knowledge management systems as the next state-of-the-art innovation pertinent to business practitioners (Adams & Lamont, 2003). For example, researchers such as Davenport & Prusak (1988), Johnson (1988), Zack (1999), and Alavi & Leidner (2001) emphasize the criticality associated with corporations developing organizational-wide KMSs to create and maintain competitive advantages in increasingly dynamic business environments (Adams & Lamont, 2003).

A number of organizations have implemented KMSs only to find that employees do not use them (Hansen & Von, 2001). Issues such as motivating employees to share knowledge (Wasko & Faraj, 2005), creating positive attitudes around knowledge sharing (Bock, Zmud, Kim, & Lee, 2005), and trust (McEvily, Perronne, & Zaheer, 2003) continue to be addressed in research and in practice. As with any other information system implementation, the success of these systems inevitably begins with the individual; individual acceptance and usage are critical (Money & Turner, 2004). With continuing business resource investments, understanding and creating conditions under which information systems will be accepted and used in human organizations remains a high priority within the research community (Vankatesh & Davis, 2000). However, understanding why individuals accept or reject systems has proven to be one of the most challenging issues in information systems research (Doll, Hendrickson, & Xiandong, 1998).
User acceptance of information systems and usage are unquestionably crucial factors in the ultimate determination of information systems success, because information systems that are not used are of little value (Mathieson, Peacock, & Chin, 2001). Similarly, creating knowledge management systems likely to be accepted by target users is critical to harnessing a new system’s potential (Lin, Hu, Chen, & Schroeder, 2004). For present purposes, user acceptance is defined as the demonstrable willingness within a user to employ information technology for the tasks it is designed to support (Dillon & Morris, 1996). User participation in system design is seen as a key factor to achieving acceptance (Mathieson, 1991). Many believe that systems developed with user participation will better match user requirements and capabilities than systems designed solely by information system professionals (Mathieson, 1991). In addition, Ambrosio (2000) asserts that the most common error in implementing systems is failing to coordinate efforts between information technology and human resources. In his literature review of information system failure factors, Malhotra (2004) noted that systems should ensure that adaptation and innovation of business performance outcomes occurs in alignment with changing dynamics of the business environment. Armed with the knowledge of why people resist using information systems, researchers can develop better methods for designing technology, for evaluating systems, and for predicting how users will respond to new technology (Gould, Boies, & Lewis, 1991). As organizations become more dependent on information systems and their use spreads across society, the concern for developing information systems that will be used becomes even more important.

Although KMSs have become a popular focus in many firms, many KMS initiatives fail to achieve their goals. There have even been major failures documented within the law-enforcement domain. Eggen and Witte (2006) describe the FBI-contracted development of a network system (Virtual Case File) for tracking criminal cases. After spending $170 million, the FBI still had an archaic computer system and had to restart development. “The collapse of the attempt to remake the FBI’s filing system stemmed from the new system being incomplete, inadequate, and so poorly designed that it would be essentially unusable under real world conditions” (pg. A01). In addition, the system could not properly sort data and lacked common features, such as bookmarking that would help agents navigate through millions of files. As a result, the FBI found the system so incomplete and unusable that they discarded the system altogether.

**KNOWLEDGE MANAGEMENT IN LAW ENFORCEMENT**

One context in which we find evidence of the need for effective and widely accepted knowledge management systems is in the discipline of digital forensics in law enforcement agencies. Law enforcement agencies possess a large but unstructured community memory with respect to digital forensics because there is not an explicit mechanism for disseminating the experiences of every digital forensic technician and investigator (Harrison, Aucsmith, Heuston, Mocas, Morrissey, & Russelle, 2002). The explosive growth in the digital information maintained in the management systems of law enforcement agencies and the spiraling need for cross-agency access to that information have made utilizing such information both increasingly urgent and increasingly difficult (Hu, Lin, & Chen, 2005).

Incompatible content and information formats often create barriers to data access and utilization that make knowledge management a complex and daunting process (Jones & Jordan, 1998). For example, information and knowledge are captured within law enforcement agencies in various forms ranging from computer records to documented institutional orders to the personal experience of digital forensic officers (Luen & Al-Hawamdeh, 2001). The crux of the issue for law enforcement is how to surface such knowledge and bring it to bear on the problems faced by digital forensic examiners in a timely and effective manner.

Digital forensic investigators also need timely access to relevant and accurate knowledge presented in an integrated and easily analyzed manner. According to Hauck and Chen (1999), the ideal knowledge management system for law enforcement agencies should be able to provide information about problems that have not been identified previously, and thus be able to give innovative and creative support for new investigations. In the case of digital forensics, the data may be available but not in a form that makes them useful for higher level processing (Hauck, 1999). For example, digital forensic investigators devise tactics, techniques, and practices that are difficult to search and analyze. Often, only experienced and knowledgeable investigators are able to use such organizational resources effectively.

There are a number of available systems that currently serve as information management or intelligence analysis tools for law enforcement (Chen, Schroeder, Hauck, Ridgeway, Atabakhsh, Gupta, Boorman, Rasmusse, & Clements, 2002). Each of these systems has its own drawbacks and implements only a certain aspect of storing and disseminating knowledge for law enforcement. Harrison et al. (2002) proposed a prototype web-based repository (Lessons Learned Repository for Computer Forensics) for sharing information, but the effort was not
widely accepted (Biros, Weiser, & Mosier, 2006) by a significant portion of the law enforcement community in a manner that allows previous discoveries to be applied to future cases.

NEED FOR A NATIONAL REPOSITORY OF DIGITAL FORENSIC INTELLIGENCE

The National Repository of Digital Forensic Intelligence (NRDFI) was designed to address the knowledge management issues across many law enforcement and intelligence agencies through an integrated system that allows investigators to access and share information with other agencies. The NRDFI, a digital forensic knowledge repository development project between Oklahoma State University’s Center for Telecommunications and Network Security and the Defense Cyber Crimes Center, is a mechanism that provides flexible information sharing between law enforcement agencies. The NRDFI aims to reduce the time required to analyze evidence and advance the investigation of current cases by capturing and correlating digital forensic intelligence related information in social and organizational contexts.

Many issues and obstacles must be addressed to ensure the successful deployment of the NRDFI in the digital forensics community. There is a great sense of ownership by law enforcement agencies and individual investigators which impacts trusts and willingness to share information and creates a kind of competition between the groups (Biros et al., 2006). There are often technical and bureaucratic barriers between various law enforcement systems. The inability to integrate and access the vast number of law enforcement management systems and the inability to share information with other systems prevents an agency from receiving timely information from other data sources ultimately decreasing the efficiency of crime prevention and investigations (Hauck, 1999).

Law enforcement professionals, and more specifically digital forensic investigators, like computer network security experts tend to rely more on personal social networks or ego-centric networks rather than more formal repositories of information thus impeding information sharing in this domain (Jarvenpaa & Majchrzak, 2005). Security and confidentiality of an investigation are additional confounds to open sharing, because inappropriate controls could lead to severe consequences. Examiners are trained to search for proven techniques which provide immediate benefit for the time invested. Because there is no immediate gain in providing information for others and a very real fear that current and future criminals may improve their own skills with this knowledge, agencies are not motivated to share.

The NRDFI project was implemented to address some of the major issues described above and mimic the way digital forensic experts work. The NRDFI is designed to allow geographically diverse law enforcement agencies to share digital forensic information that will hopefully aid every agency in successfully prosecuting their case (Biros et al., 2006). In its full implementation, the NRDFI has the potential to provide exceptional gains in efficiency for forensic examiners and investigators by providing a better conduit to share relevant information between agencies and a structure through which cases can be cross referenced to have the most impact on any current investigation (Biros et al., 2006).

NRDFI OVERVIEW

Details of the early NRDFI and its underlying design can be found in (Biros, et al, 2006). Based on feedback from over a year of active use at DC3, interface and features have been redesigned for deployment in multiple beta agencies within the Department of Defense and law enforcement. This paper provides a brief description of the overall design, as well as new features that have been implemented to address many issues raised in earlier research.

The essence of the repository is to capture and share the best practices of examiners with those who would otherwise need to discover or develop the same or similar techniques. The types of documentation and information contained therein are widely varied. A common search mechanism across all information is critical for finding earlier discoveries that can be brought to bear on current cases. The social networks that are heavily relied upon in law enforcement have driven the underlying structure of the NRDFI.

Each agency has its own repository in which it can maintain its own information in a very flexible structure that can be adapted to best match that organization’s methods. The repository supports virtually any type of binary or text files and we continue to add parsing mechanisms for document types that will allow full-text searching of
submitted items. We recognize the wealth of information that is publicly available on the Web, so we also support any URL-addressable item as a resource and will parse that for searching as well.

Resources can be grouped into panes that make sense in the agency. A subset of DC3’s repository is shown in Figure 1, with groupings that align to their internal agencies and other commonly used items within the group. Panes or individual documents can be shared with any agency or group that the administrator has allowed. Individual users can also customize their own screen by moving or hiding panes that are available to them.

![Figure 1: Agency Repository Interface](image)

**COMMUNICATION SUPPORT**

A need exists for a common secure communication mechanism across law enforcement agencies. Among other methods, agencies encrypt and e-mail a document over standard e-mail systems and then call the recipient with the password. The NRDFI provides two mechanisms to support online discussion. The first is a threaded discussion forum that can be created by any user. It can be attached to any resource or pane, or be created as a stand-alone resource. As a resource, it can also be shared with other cooperating agencies and appropriately vetted users in the same manner as any other document.

There is also a secure communication mechanism that serves as a secure messaging system. It has significant additional flexibility over the function of the rest of the NRDFI. Any user with access to this feature can communicate through the system with any other user on any repository, regardless of whether or not that person’s agency is allowed to receive other resources. Communications are for groups of users, rather than be
limited to two participants, and the membership is controlled by the person who initially creates the communication. Unlike all other sharing capabilities, the secure communication feature can extend to users who are not vetted in any repository. There is a separate secure messaging server into which any user can invite any person with HTTPS access and a basic e-mail account. Once that user creates an account, they are admitted into one or more communications to which they have been invited.

Secure messaging also has two different structures. The default mechanism simply displays interactions in reverse chronological order, allowing participants to view the entire history of the communication. If a document is being revised collaboratively, the second mechanism allows a wiki-like structure, where a common document is edited and marked up by all invited users in an asynchronous approach.

ACCESS RESTRICTIONS

Users who have access to the information within that agency are vetted by a local administrator who can also assign a certain classification sub-level. The following levels are offered as a default for both users and documents. They are strictly hierarchical and additional levels can be added in an agency for refinements when that is found to be necessary.

Unclassified – Law Enforcement Sensitive: Are particularly sensitive, especially in pending cases and could be damaging to current or future cases if the information were made available beyond the law enforcement community

Unclassified – For Official Use Only: Should not be made available outside the agency, unless there is a specific reason to do so

Unclassified – General: Available to everyone who has an account on that repository or another repository that has read access

Any user granted access at a certain level has access to documents at that level and below. Additionally, the classified hierarchy can be used on a network that is certified for that type of information. If an agency chooses to insert additional levels between these, they too are strictly hierarchical.

INTER-AGENCY SHARING

Inter-agency sharing is made possible through a mesh structure in which every repository administrator selects peer repositories with which they choose to cooperate (thus mimicking the ego-centric networks of digital forensic specialists). By default, there is a core repository to which everyone has read access and administered posting access. Like any repository in the system, however, an administrator may choose to not allow his local users to access information from the core.

Grouping of outside repositories is also possible. For instance, a Department of Defense group may be established. When a user wishes to share a document to all DOD agencies that have been allowed by the administrator, he can simply select that grouping, rather than individually selecting each agency. That also facilitates future DOD agency repositories’ immediate access to all these shared documents.
Figure 2: Flexible Mesh of Repositories

Figure 2 shows several different examples of how this may be implemented. The diagram is not intended to reflect current or planned cooperative relationships between specific agencies that might participate in the repository. It is provided purely as a notional illustration:

DC3 has a repository for storing information that they want to make available to their investigative agencies, but not outside the DOD, although the Naval Criminal Investigative Service (NCIS), the Air Force Office of Special Investigation (AFOSI), and their field offices can directly use and contribute to the core repository as well, or retain data only within their agency without elevating it even to the level of DOD.

The FBI offices have a similar structure, but one of the field offices may cooperate extensively with one of the NCIS or AFOSI field offices in the same city and liberally share new discoveries with each other. This creates a new “neighborhood” that is labeled local coop in the figure.

Small agencies may have a single repository for their lessons learned, but they share with the core repository. In the extreme, there may be no local storage at all, but a web interface directly into the core repository. A small sheriff’s office with a forensic capability can leverage the lessons learned in many other participating agencies with little investment.

Some data is very sensitive. In the figure, the NSA is shown with a neighborhood among its own central node and field offices, but only as a consumer of data from the central repository. This will not benefit other agencies; however, some organization’s requirements will prohibit sharing information.

Some agencies will choose to be entirely isolated. They can neither benefit from the central repository nor enhance it, because of a logical and/or physical separation. The underlying system design, however, allows them to share among their own neighborhood, while retaining complete control of hardware, software, and data.

Because a DOD grouping has been established with all relevant agencies, when a document is shared that is intended only for people vetted in those repositories, the user need only select that group, rather than all the repositories shown in the bottom oval. This does not override sharing relationships within the oval, so the intersection of DOD agencies that have sharing from the source agency will see the document.
FUTURE RESEARCH

We believe that many systems fail even when they achieve the TAM goals of Usability and Ease of Use. Through our beta test and a series of data collection methods, we hope to demonstrate that developing an effective NRDFI requires development in accordance with the way digital forensic examiners work. First, we will collect data on the nature of digital forensic work from a large body of specialists at a national conference in early 2008. Second, we will use structured interviews to better understand the needs of the examiners. When complete, we will then use the data to refine the NRDFI to better meet the needs of the examiners.

CONCLUSION:

The increase in digital forensic cases far outpaces the growth in numbers of forensic examiners. General requirements for legal admissibility, however, are strict and unchanging. With constant modifications to the technologies that are examined, mechanisms to share new knowledge are critical in keeping up. An information repository that allows geographic and bureaucratic agencies responsible for the analysis to communicate and share new discoveries may be the only way to efficiently and effectively process these cases.

The collaborative effort between Oklahoma State University and the Defense Cyber Crime Center seeks to fill this need. The NRDFI prototype implementation has provided a great deal of feedback about how to overcome impediments that have been recognized in prior research. Technical constraints and relationships between repositories, as well as limitations on document access attempt to support the social networking that is currently used between agencies and personnel. The wealth of information that will be available in a widely-adopted system on this platform will be invaluable to assist investigators who are working with unfamiliar cases.

Although we have addressed many of the recognized impediments, there are still social issues that will prevent some individuals and agencies from sharing. By augmenting the system with private communication mechanisms, we hope to address the needs of even the most conservative examiner. New examiners, however, should be able to come up to speed and serve their constituents much more quickly with the wealth of knowledge that would immediately be made available through their agency’s and cooperating agencies’ repositories. Both the experienced and novice agent, however, will better be able to leverage the knowledge of others for successful legal outcomes.

REFERENCES:


COPYRIGHT

Biros, Weiser, Whitfield ©2007. The author/s assign Edith Cowan University a non-exclusive license to use this document for personal use provided that the article is used in full and this copyright statement is reproduced. Such documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. The authors also grant a non-exclusive license to ECU to publish this document in full in the Conference Proceedings. Any other usage is prohibited without the express permission of the authors.