This document is an author-formatted work. The definitive version for citation appears as:

R. C. Watkins, K. M. Reynolds, R. F. DeMara, M. Georgiopoulos, A. J. Gonzalez, and R. Eaglin, "Exploring Data Mining Technologies as Tools to Investigate Money Laundering," *Journal of Policing Practice and Research: An International Journal*, Vol. 4, No. 2, January, 2003, pp. 163 – 178. ISSN 1561-4263 print/ISSN 1477-271X online/03/020163–16 © 2003 Taylor & Francis Ltd. DOI: 10.1080/1561426032000084936

Tracking Dirty Proceeds: Exploring Data Mining Technologies As Tools To Investigate Money Laundering

R. C. Watkins, K. M. Reynolds, R. F. DeMara, M. Georgiopoulos, A. J. Gonzalez, and R. Eaglin
University of Central Florida,
4000 Central Florida Blvd,
PO Box 161600
Orlando, FL 32816–1600, USA

Money laundering enforcement operations in the USA and abroad have developed in the past decade from the simple use of informant information to the sophisticated analysis of voluminous, complex financial transaction arrays. Traditional investigative techniques aimed at uncovering patterns consume numerous man-hours. The volume of these records and the complexity of the relationships call for innovative techniques that can aid financial investigators in generating timely, accurate leads. Data mining techniques are well suited for identifying trends and patterns in large data sets often comprised of hundreds or even thousands of complex hidden relationships. This paper explores the use of innovative data mining methodologies that could enhance law enforcement's ability to detect, reduce, and prevent money laundering activities. This paper provides an overview of the money laundering problem in the USA and overseas and describes the nature and scope of the money laundering problems. It reviews traditional approaches to financial crime investigation and discusses various innovative data mining and artificial intelligence-based solutions that can assist financial investigators.

Keywords: Money laundering; Financial crime investigation; Artificial intelligence; Neural networks; Information technology

INTRODUCTION

In recent years, law enforcement and the federal government have embarked on a vigorous mission to investigate the laundering of illicit proceeds gained from drug trafficking activities. Substantial amounts of illegally obtained funds are laundered through businesses in the USA and overseas. Estimates suggest that US$500 billion to US$1 trillion is laundered annually worldwide (Scott, 1995; Robinson, 1996; Baker, 1999). Financial crime, and in particular the laundering of proceeds acquired through illegal activities, effects any country that is integrated into the international financial arena.

The investigation of money laundering violations began as a federal effort under the authority of the provisions of the US Code that prohibits attempts to conceal the illicit origins of crime profits (see Money Laundering and Financial Crimes Act of 1998). Moreover, international agreements have been struck to address this pervasive and harmful form of criminal activity. Some examples of multi-national agreements aimed at reducing money laundering activities are the International Organization of Securities Commissions (1992), United Nations Vienna Convention against Illicit Traffic in Narcotic Drugs and Psychotropic Substances (1990), the Council of Europe Convention on Laundering, Search, Seizure, and Confiscation of the Proceeds of Crime (1990), the Financial Action Task Force (1989) and the Basle Commission on Banking Supervision (1988).

The State of Florida passed criminal laws designed to address the large volume of money laundering transactions occurring in the Miami area that were connected to narcotics trafficking. Whereas early enforcement initiatives were led by federal agencies from the US Departments of Justice and Treasury, a state and local money laundering task force was formed in 1994 under the auspices of the federal South Florida High Intensity Drug Trafficking Area (HIDTA) partnership. This enforcement effort has led to the confiscation of over US$100 million in cash proceeds and to the seizure of large quantities of narcotics. The success of various federal, state, and local efforts has led to the recent commitment of significant additional investigative resources to the problem of concealing illicit funds. The investigation of money laundering activities is a complex and intensive resource endeavor.

Banking institutions are required to complete a currency transaction report (CTR) and suspicious activity report (SAR) for all deposits, withdrawals, and currency exchanges over US$10,000. Many of these transaction reports are completed and submitted on a daily, weekly, and monthly basis. For example, one New York bank handles approximately 40,000 wire transfers each day (Abadinsky, 1997). At the federal level, it is estimated that the government receives more than 7 million CTRs each year. Furthermore, the average daily volume for American transactions is approximately US$7 billion, and on one particular day, it was estimated that US$1.25 trillion was processed through the banking industry (Ratledge and
Jacoby, 1989).

Uncovering instances of money laundering and subsequently linking transactions to drug trafficking poses a serious challenge to law enforcement. Currently, investigative leads are developed through resource-intensive, manual scrutiny of centralized repositories of financial data. Law enforcement could benefit significantly from the use of new technology designed to automate the analysis of financial transaction data. The rapid identification of money laundering patterns through automated analyses can more efficiently and effectively produce investigative leads to support enhanced money laundering and drug enforcement operations.

This paper explores the use of new computer technologies that could assist law enforcement identify patterns of money laundering to assist them in their efforts to reduce financial and drug-related crimes. We break this paper into four sections. Section one describes the nature and scope of the money laundering problem, paying close attention to the process by which money laundering is performed. Section two identifies traditional investigative approaches for combating money laundering. The third section is devoted to an exploration of innovative technologies that could be used to assist law enforcement suppress financial-related crime more efficiently. In this section, we describe each methodology and provide examples of each application's utility for financial investigators. We conclude this paper with a discussion of the benefits and limitations of data mining technologies, and those factors to be considered by financial investigators when integrating these applications into their repertoire of tools to combat money laundering.

NATURE AND SCOPE OF THE PROBLEM

Money laundering has been defined as ". . . to knowingly engage in a financial transaction with the proceeds of some unlawful activity with the intent of promoting or carrying on that unlawful activity or to conceal or disguise the nature, location, source, ownership, or control of these proceeds" (Genzman, 1997: 342). The money laundering process is often used by drug dealers to disguise their illegal profits. These illicit drug profits are laundered by converting them into other legal assets, such as real estate, stocks, gold or other valuables (Schmallenger, 1999). According to the National Narcotics Intelligence Consumers Committee (NNIC) report of 1996, millions of dollars in drug monies are laundered through commercial banks and other financial institutions each year, with major money laundering operations flourishing in South Florida and Los Angeles.

The process of money laundering has been a long-term nemesis of law enforcement. The complexities of money laundering techniques have increased dramatically, thereby making it difficult for law enforcement interventions. Advances in banking technology make it particularly difficult for law enforcement to effectively monitor banking transactions.

Intermingled with these obstacles are the massive volume of wire transfers handled by banks on a daily, weekly and monthly basis. It was determined that one New York bank handles approximately 40,000 wire transfers per day (Abadinsky, 1997). Banking institutions are required to complete a CTR for all deposits, withdrawals, and currency exchanges greater than US$10,000; however, the government receives more than 7 million of these reports each year and is heavily backlogged (Abadinsky, 1997). Phil Williams, Director of the Ridgeway Center for International Security Studies states, "It took 45 seconds to launder the money by wire transfer, and it took the police officers 18 months to investigate the case" (2000 online).

The sheer volume of these reports impedes law enforcement's ability to efficiently investigate these cases. Andrew Haynes suggests the main effect of tracking CTRs has been "to fill warehouse space in Detroit with paperwork to little measurable benefit from the point of view of law enforcement" (1996 online). The complexity of the money laundering process is designed to protect the anonymity of those involved in illicit activities. In order to prevent detection from law enforcement, several steps must be taken to ensure the money is nearly impossible to trace back to its original source. The money laundering process can be divided into three distinct phases: placement, layering, and integration (Madinger and Zalopany, 1999).

Placement

The first phase, and arguably the most critical, is placement. At this point the criminal enterprise has yielded cash profits. Large amounts of cash must be converted from currency into some other form. It must be moved from a cash-based transaction system to a businessbased transaction system (Madinger and Zalopany, 1999). This usually involves placing the money in a financial institution. Large amounts of money can be transformed into cashier's checks and then placed in a bank, as was done in the Watergate laundering of illegal corporate campaign contributions. The criminal is very vulnerable to compromise at this point because financial institutions tend to ask questions and file reports regarding large transactions. Thus, the money must be divided into less conspicuous amounts and placed, preferably, into a wide variety of financial institutions. If this transaction goes smoothly, the money laundering process will proceed to the next step referred to as layering.

Layering
During the layering stage, the money that had been placed is moved from institution to institution. This is routinely done to make it more difficult for investigators to construct a paper trail. According to the Financial Crimes Enforcement Network (FINCEN):
Money is still smuggled out of (or into) the United States in the form of currency, but law enforcement agencies expend great resources trying to stop criminals from physically smuggling their cash profits across national borders, only to have the money flow without hindrance through electronic communication systems to countries where bank accounts are protected by secrecy laws. (Senator, 1995: 21)

It is during the layering stage that an effort is made by criminals to ensure that laundered money cannot be differentiated from funds derived from legal activities. This is done by exploiting the 'frequency, volume, or complexity' of transactions. The process of layering encompasses the assurance that the transactions cross several national borders. This "transjurisdictional" money movement is a key element of most money laundering schemes (Williams, 2000). Once the laundered money can no longer be differentiated from money obtained through legal activities, money launderers can then begin the final phase of the monetary cleansing process: the integration phase.

Integration
It is during this final phase that criminals attempt to integrate illegally obtained proceeds into legitimate businesses. Often cash-intensive businesses (e.g., nightclubs, bars, restaurants) are used as fronts because it is somewhat easier to over-report earnings (Standberg, 1997). Once the legitimate businesses are established, it becomes easier to launder funds using a variety of methods. First, a business may overstate the amount of revenue it is receiving through normal operations. A second method is to overstate the expenses required in the operation of the business. Still, a third approach is to write checks and deposit cash in excess of both reported revenues and expenses (Zeldin, 1989). The steps associated with money laundering – placement, layering, integration – are a universal process that effects financial investigators within and outside the borders of the USA. As noted by the Financial Action Task Force and its constituent members, no new methods of money laundering have been identified (FATF Money Laundering Public Report 1997). The criminal first places the funds in an offshore account by removing them from the USA through the use of currency exchanges or electronic wire transfers. The second step is to find a means of legitimizing the funds. The final step in international operations is to repatriate the funds without detection from the Internal Revenue Service (Webster and McCampbell, 1992). This three-step process characterizes how money is laundered in the USA and overseas.

The money laundering process is extremely complex and requires a tremendous resource base to expeditiously identify accurate, high quality leads that can inform financial crime investigators of suspicious banking practices. Presently, financial investigators worldwide lack the necessary resources to examine large financial data sets. Consequently, financial crime investigators are routinely relegated to traditional approaches to combating the money laundering problem in their respective jurisdictions.

TRADITIONAL APPROACHES TO COMBATING THE MONEY LAUNDERING PROBLEM

Uncovering instances of money laundering poses a serious challenge to law enforcement officers. To identify patterns of money laundering requires substantial training in the area of accounting

and banking operations; neither of which are emphasized in law enforcement training academies. Most law enforcement officers in general are trained to deal with the perpetrator of street crimes. White-collar crime is usually left to federal agents who either possess an expertise in accounting and banking or have received extensive training in these areas. These trained professionals are the only individuals qualified to peruse the records and account ledgers of suspected money launderers.

Manpower is another important issue that impedes law enforcement's ability to track money laundering activities. There are simply not enough federal agents to track down more than a fraction of the billions of dollars laundered annually. Moreover, on top of the insufficient training and manpower deficits that routinely plague financial crime investigations, there is the problem of knowing where to look. Agents are tasked with identifying the money launderers' transactions amidst millions of other financial transactions. These are the most common problems that impede police operations and investigations of illicit financial activity. Consequently, traditional investigative methodologies and techniques are inefficient and resource intensive. Traditional investigative approaches used to uncover money laundering patterns can be broken down into one of three categories in an overall cumulative process: identification of money laundering incidences, detection avoidance, and surveillance of money laundering activities.

IDENTIFICATION OF MONEY LAUNDERING INCIDENCES

The first step in launching an investigation of this sort is knowing that the business or individual to be examined is suspected of illicit activity. A tip might come from a disgruntled employee, or a bank can become suspicious by the high volume of activity done in a business' or individual's name. Another source of information may come from informants, who may provide tips concerning criminal activity and/or financial transactions (Madinger and Zalopany, 1999).

Traditionally, informants have been an important source of information in the investigation of money laundering. A case against a suspect is made significantly stronger if the investigator can find someone who was involved in, or at least aware of, the operation of illegal business practices. There are, however, risks inherent in using informants. The investigator should be aware that the informant is not permitted to do anything the law officer cannot do legally. The law has recently become very specific about how informants should be handled. It is also possible that the use of an informant may tip off a suspect. Regardless of how an investigating unit becomes aware of possible criminal activity, many traditional and risky investigative techniques might be employed to gather additional intelligence.

Detection Avoidance

After an investigation begins, the risk of detection can increase depending on the traditional approach used. Investigative techniques can range from low-risk computer database analysis to a much riskier approach: interviews with a suspect's associates, or even the subject under scrutiny (Madinger and Zalopany, 1999). Figure 1 identifies the array of traditional investigative techniques employed in a particular money laundering investigation. This chart also depicts the risk vs. reward structure based upon the investigative technique selected. The least risky technique used in financial investigations is the analysis of computer databases. This can be highly effective in lead generation, however, it is extremely time consuming given the limited training and manpower shortages of many police agencies. Investigators do, however, have access to numerous financial data sets. For example, the Treasury Enforcement Communications System, or TECS, gathers and stores millions of records, such as CTRs, from many government agencies. Agencies such as the US Customs Service, US Secret Service, the Internal Revenue

Service, and the Bureau of Alcohol Tobacco and Firearms are linked to TECS, allowing those organizations to store and retrieve various documents. Other computer databases include the Narcotics and Dangerous Drugs Information System, or NADDIS, the National Law Enforcement Telecommunications System, or NLETS, and FinCEN's Artificial Intelligence System, or FAIS. Unfortunately, procurement of these data often takes a long time. Furthermore, it also takes substantial time to process the data and isolate patterns and leads. Additionally, access by non-federal law enforcement agencies may be restricted (Madinger and Zalopany, 1999). These three factors are perhaps the most significant obstacles to successful money laundering investigations as the inability to identify accurate, timely leads commonly results in a stalled or failed investigation.

Surveillance of Money Laundering Incidences

Surveillance operations encompass yet another umbrella of traditional investigative techniques aimed at gathering intelligence on a suspect. There are many forms of surveillance that are used to uncover illicit financial activity. These range from the least intrusive methods such as discrete inquiries with potential suspects or associates, to mail and trash covers, to much more risky, intrusive techniques like suspect interviews. While surveillance is not a new technique, recent advances in technology can make surveillance easier for investigators (Madinger and Zalopany, 1999). Video and audio advances can decrease the probability that an investigative team will be detected. It can also increase the amount of information gathered. One form of surveillance, mail or trash covers, can reveal various financial transactions. To examine mail, a written request of the Postal Inspector in charge is necessary.

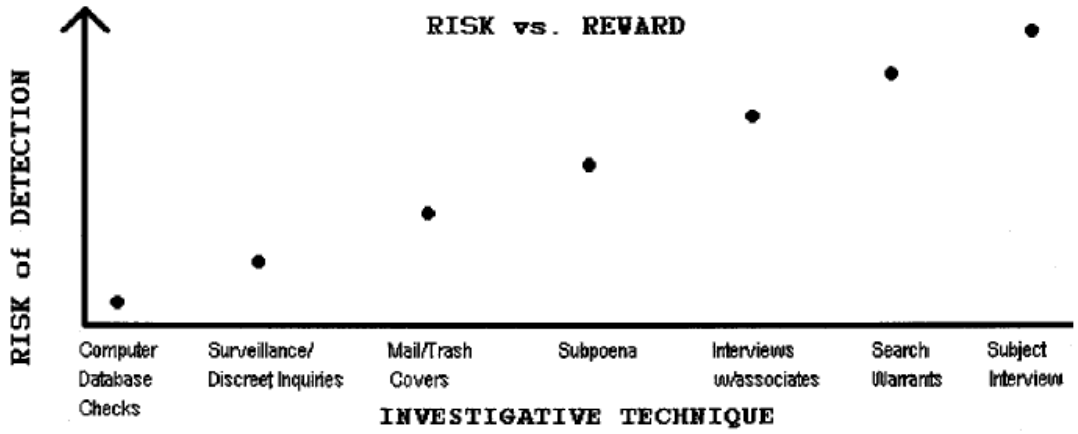

Figure 1: Investigative technique risk reward structure.

This will provide information on all known addresses of the suspect or business. This can make it easier to determine if the suspect was living beyond his or her means (Nossen and Norvelle, 1996). The mail cover will also enable the investigator to identify the location from which a suspect's mail is being generated. Even the postmark itself on the mail is sometimes important. It is helpful to know if the suspect is receiving mail from the Cayman Islands or Zurich (Madinger and Zalopany, 1999). Surveillance information generated from trash or mail covers is beneficial to financial investigators because while it is illegal to actually open the mail, merely observing who is sending the suspect mail is often very useful. Banks, brokerage houses, and other businesses working with the suspect or his business can also be tracked down as a result of careful surveillance.

Irrespective of the method by which money laundering intelligence is received, financial investigators need to be aware that businesses, such as banks, do not readily turn over their records of client transactions, even to law enforcement. Financial investigators must file the proper paperwork to gain access to the bank's records and, ideally, the investigator should have a subpoena for the necessary information. If a subpoena cannot be readily obtained, it may be easier to concentrate on the aspects of the business that are public domain. If the business is large enough, it will have an annual report. The annual report tells the investors and the Securities and Exchange Commission how much money the business earned and how it is spending its income. This, once again, reinforces the importance of proper training for those investigators assigned to uncovering money laundering activities.

Proceeding through the investigative process thus far, an investigating agency may have gathered enough facts to confirm suspicions about a particular suspect or business. A subpoena or perhaps a search warrant would be justified. However, more evidence may be required to ensure that the proper charges are filed. An investigator might move towards a more risky form of surveillance such as conducting an interview with the suspect's neighbors, co-workers, and other associates to gather additional information. An investigator may even attempt to interview the suspect, though this requires careful consideration. Often, premature disclosure results in a frenzied effort by the suspect to purge records and accounts. Also, the suspect may choose not to comply with any questioning. At this point, a search warrant would have to be requested to allow the investigating agency to analyze the subject's records.

This sort of investigation consumes substantial time and resources. It is not uncommon for federal officials to work a specific case for a number of years. Alternatively, the individuals responsible for laundering illegal monies accomplish their mission and do not need nearly that much time to launder their illegal profits. Once again, law enforcement is severely restricted by their limited manpower and training in this area. The outcome is usually financial crime investigations that continue well after the illicit proceeds have been successfully laundered. In the USA there are only 10 small groups working for the FBI that specifically handle money laundering. Since most money laundering cases cross state lines, and often cross international borders, local law enforcement usually hands these cases over to an already overworked group of understaffed federal agents.

To resolve some of the obstacles faced by law enforcement in the investigation of financial crimes, it is proposed that a number of innovative technological solutions exist. For instance, the use of a variety of data mining techniques could significantly decrease the time it takes investigators to uncover case leads. These techniques can also generate new leads that perhaps go unnoticed by subject experts that have understaffed teams of investigators and are burdened with manual examination of large financial data arrays.

NEW INNOVATIVE APPROACHES TO COMBATING MONEY LAUNDERING

Data mining refers to the collection of automated tools and Artificial Intelligence (AI) techniques that facilitate searching of large data sets to discover 'hidden' or 'buried' relationships among the data, not easily identified by inspection. Data mining concepts have been evolving for decades. But, it was the emergence of the data warehouse, increased computer processing speeds, and enhanced graphical user interface (GUI) software packages that have made data mining tools accessible to the user community and very well suited to assist in the investigation of money laundering activities.

A data warehouse is a single integrated database where 'scrubbed' data from a variety of sources exist. So, where traditional databases relied on the end-user to recognize problems or opportunities, select relevant data and write queries and supply answers, data mining tools transform raw data into information that can be processed into knowledge by the end-user. To effectively use data mining techniques, the data mining software must have access to the information in a usable and cogent form. To provide this information, the data is transferred and stored in a centralized data warehouse. The data warehouse is more than just a storage location – it also contains tools to allow data import, extraction, analysis, and reports. To increase effectiveness, the warehouse incorporates into its interface tools that can be used by the end-user to allow data mining to be performed natively within the database system. A significant benefit of data mining methodologies is that they can lead financial investigators to the discovery of previously unknown information that may prove extremely valuable to the investigation.

One of the initial steps in the investigation of suspected money laundering activities is the analysis of computer databases. This is critical to the investigation because it is at this point that investigators are expected to develop timely leads to cases. Unfortunately, as a consequence of restricted manpower, limited training, cost, and out-dated computer systems, financial investigators are forced to manually comb large financial data sets to identify laundering patterns and generate leads for case follow-up. This is very resource intensive and often results in information that is incomplete and untimely. It also results in failures to identify substantive case leads. It is at this point where state-of-the-art data mining tools could be of substantial assistance to all individuals engaged in financial investigations. Because these tools are designed to facilitate automated searches of large data sets and rapidly discover "hidden" or "buried" relationships among the data, they can be extremely valuable to financial investigators who have traditionally searched voluminous data sets manually.

THE APPLICATION OF DATA MINING TECHNIQUES TO MONEY LAUNDERING

The current level of sophistication in applying AI tools to money laundering is rapidly increasing. Senator (1995) describes the specifics of an AI tool, named Financial Crimes Enforcement Network AI Systems (FAIS). This data mining methodology evaluates reports of large cash transactions (CTRs and SARs) to identify potential money laundering activities and patterns. FAIS uses data fields obtained from the CTRs to conduct evaluations of suspicious financial activities. FAIS eventually uses 336 rules to determine the suspiciousness of a report. Each rule contributes positive or negative evidence as to whether the transaction, subject, or account is suspicious or legitimate. The evidence from each rule is combined to derive a single suspiciousness rating for the transaction, subject, or account. While the FAIS tool does have clear application in assisting with money laundering problems, it has the limitations of an expert system. Essentially, its performance is affected by the levels of expertise that are readily codified by the investigators and domain specialists. Due to the complex nature of money laundering, it is challenging to derive a single rule set capable of explicitly expressing all of the relationships in the data that correspond to illicit activities. Financial investigators worldwide continuously struggle to identify a robust set of rules that encompass the diverse range of money laundering activities. While these efforts take time and resources, technological alternatives do exist that could aid investigators in refining rule sets and identifying case leads. One unique benefit of AI technology is that, while it is initially restricted by the expertise levels of domain specialists, these technologies do have the capacity to learn and identify previously unknown patterns and linkages. This, of course, depends on the quality and quantity of the data set(s) used to train the

models. Nonetheless, these technological solutions have the ability to save financial investigators time and improve investigations.

Louie and Kraay (1999) identified advantages of one data mining tool used by law enforcement: the Origami software package. They also listed a number of applications where Origami could be used in law enforcement. One application involves the mining of telephone calls made by individuals that might reveal illegal activities to investigating officers. For example, if phone number A calls phone number B, and then phone number B immediately calls phone number C, Origami establishes a virtual link between phone numbers A and C. While this example is simplistic, network analyses can rapidly and accurately be produced using data mining tools.
The *Handbook on artificial intelligence and expert systems in law enforcement*
(Ratledge and Jacoby, 1989) emphasizes the importance of AI tools helping law enforcement attain their goals. In the foreword of this book, Bentham (1998: 342) states ". . . in the 1990's, law enforcement must show leadership and foresight in planning for and adopting the next generation of automation-computer assisted investigation and decision making." Moreover, Piatesky-Shapiro (1996) provide an overview of issues associated with the development of powerful data mining and knowledge discovery applications. They identify three types of data mining tools: the generic single-task tool, the generic multi-task tool, and the domain-specific tool. The domain-specific tool supports discovery in an explicit domain and supports the language of the end-user. For the money laundering application, it is most beneficial to utilize a domain-specific tool. Piatesky- Shapiro (1996) offer a list of applications in marketing investment and fraud detection where data mining tools have been useful. Furthermore, Eick and Fyock (1996) describe a system that was developed for detecting international calling fraud while Davis and Goyal (1993) discuss the development of a system where customer profiles are used to detect cellular cloning fraud.

There is emerging interest in the research and practitioner field concerning the use of AI tools to enhance public safety. Trippi and Turban (1996), and Refenes (1995) offer excellent examples where applications of AI are used in finance. Furthermore, Yuanhui *et al.* (1997), Frayman and Wang (1998), and Ciesielski and Palstra (1996) also make reference and demonstrate the potential utility of applying AI techniques to the investigation of suspicious financial activities.

APPLICABLE AI TECHNIQUES AND THEIR UTILITY TO FINANCIAL INVESTIGATIONS

Some of the most popular tools in data mining applications rely on statistical and AI techniques outlined in Table I.

The most basic approach to data mining, a linear regression model, is designed by defining a dependent variable (output) and a number of independent variables (inputs). The result of a linear regression model is an equation of a line that best fits the data set, which can be used for prediction purposes. This is useful for discovering, validating, and quantifying trends from previously solved money laundering cases for use on current cases. Linear regression techniques center around the ability to predict useful quantitative probabilities. For instance, data from previously observed behaviors can be used to focus new investigative activities on the most promising locations at the most probabilistically promising day and time. Logistic regression is a very popular means of data mining because it can solve problems involving categorical variables (e.g., variables that can be described by a yes/no answer or male/female answer). This technique can be applied to rapidly evaluate all financial transaction records belonging to classes of interest to the investigator. The results can be displayed in a number of graphical formats so that

commonalties among the subset of the variables selected become evident. Using logistic regression, graphic views of only the trends in the data, rather than the data itself, are displayed in multi-dimensional format using distinguishing shapes and colors without the need for the investigator to sort through the underlying data. Consequently, investigators will be presented with improved ways of viewing large masses of data in a manageable format. This makes pattern and trend identification more apparent and timely. This approach will also serve to expedite the investigation process by reducing the amount of time spent manually searching for case leads or patterns of illicit activity.

Table I: AI techniques for detecting and combating money laundering activities

Technique	*Description*
Linear regression	Most basic approach. Predicts values by describing the linear relationship between a dependent variable and one or more independent variables.
Logistic regression	Involves categorical variables such as "yes/no" or "male/female". Very popular.
Cluster analysis	Requires substantial amounts of data that can be grouped categorically.
Inductive algorithms	Algorithms that generate decision trees based on historical outcomes.
Neural networks	An AI technique that mimics the human brain by learning from and storing inputs and outputs. Can be used with continuous/categorical variables and non-linear and collinear data.
Fuzzy logic	A theory that allows incomplete information to be processed and conclusions derived.
Genetic algorithms	Algorithms based on evolutionary rules used to solve optimization tasks.

Cluster analysis is another technique that can be used to mine large data sets for investigative lead generation. This method is applicable when one has substantial amounts of data and there is a desire to group the data into a few categories based on their similar attributes. Here the technique discovers new relationships automatically, without initial guidance from the expert. Essentially, as of yet undiscovered similarities between solved cases and the current case are enumerated. These hypotheses become powerful inputs to further refine and focus investigative efforts. Cluster analysis could be useful to investigators as it can generate additional leads for investigation, along with a deeper understanding of the geographical and chronological clusters of activity. This technique can be used to isolate statistically significant relationships between suspect networks, modes of conveyance, and locations from which drugs and illegal proceeds are exchanged. It can also function to clarify activity timelines. In other word, cluster analysis can be used to uncover similarities in the time order of events involved in suspected money laundering and drug-related investigations. Inductive algorithms is yet another AI technique that can assist national and international investigators uncover money laundering patterns. This methodology can be used to generate decision trees, and thereafter be used for the production of rules from historical occurrences of the data. This technique creates classifications of the financial transaction data by organizing it into categories, which match some pre-determined decision pattern characteristic of the activity under investigation. This is presented as a tree of decisions leading to mutually exclusive alternatives of relationships possibly indicative of money laundering. The utility of inductive algorithms stems from their ability to create a sort of 'reasoning framework.' This helps the investigator understand the functioning of a particular laundering operation. In other words, the results from this approach provide financial

investigators with an array of unique paths. These paths can then be used by investigators to understand money laundering processes and solve financial crimes. Furthermore, from these various paths, commonalities or patterns can then be established and used to (1) revise and improve old rule sets, (2) establish new rule sets, and (3) provide tightened frameworks that can assist financial investigators in money laundering investigations.

One other data mining methodology that could enhance money laundering investigations is neural networks. This technique utilizes digital computers to mimic the operation of the learning structures that exist in the human brain. A neural network accepts several inputs, performs a series of mathematical operations on the data and produces one or more outputs. The functionality resembles the human brain because (1) neural networks learn from examples of inputs and desired outputs, and (2) they store this knowledge in a distributed fashion in a neural network through a set of interconnection weights (Piateskey-Shapiro, 1996). The appeal of neural networks is they can handle continuous as well as categorical variables simultaneously, as well as non-linear and collinear data. Essentially, they can be used to learn the relevant strategies that were most helpful to the investigators so that those techniques can be given priority in the analysis of future cases. There are other AI technologies that have supported the use and optimization of the aforementioned models. Examples of these technologies are fuzzy logic and genetic algorithms. Investigators routinely handle cases where the information they have is incomplete, redundant, and imprecise. This often results in the inefficient expenditure of investigation time. Fuzzy logic and genetic algorithms can aid in reducing the amount of time required to sort through dirty data.

Fuzzy logic methodologies are well suited to deal with incomplete and imprecise information. For example, it is difficult for an investigative expert to specify a single numerical threshold value for the word ‘frequent’ in the phrase ‘frequent deposits.’ Clearly, this is a fuzzy rather than crisp value in practice. Fuzzy logic is a theory that allows users to represent incomplete and imprecise information, process it, and derive conclusions from this processing that are meaningful to investigators.

The utility of genetic algorithms also stems from their ability to save investigators time when tracking money laundering operations. This technique functions according to evolutionary rules (e.g., reproduction, mutation, crossover), and they are used to solve a variety of optimization tasks. From a financial investigator’s standpoint, such a technique could be used to delineate the most powerful leads that point toward a potential suspect or group of suspects. This can provide financial investigators with solid leads to pursue which in turn may lead to more efficient use of time and resources and lead to the closure of more cases.

BENEFITS AND LIMITATIONS OF DATA MINING TECHNOLOGIES

The Financial Action Task Force (FATF) has prescribed 40 recommendations that act as the blueprint and international standard for anti-money laundering programs. FATF has gone a long way to explicitly outline the role of financial institutions in the battle against money laundering activities. Consequently, many countries have embraced the recommendations set forth by FATF and agreed to loosen domestic bank secrecy provisions when financial investigations entail international links. This has reduced formidable barriers that have traditionally hindered money laundering investigations.

The FATF's program does not, however, focus on those obstacles and barriers encountered by line level financial investigators. These individuals are overwhelmed with voluminous data sets that are being manually examined for illicit trends and patterns. This is an incredibly time consuming and inefficient method for generating case leads and identifying money laundering patterns. Financial investigators are also burdened with insufficient manpower and technological resources.

BENEFITS TO USING DATA MINING TECHNOLOGY TO ASSIST IN MONEY LAUNDERING INVESTIGATIONS

Data mining applications offer a potential solution to the problems encountered by line level financial investigators. There are at least three areas where data mining can make a difference in financial investigations. First, incorporation of these techniques into an investigator's repertoire could reduce the amount of time spent manually combing large data sets in search of leads and patterns. Traditionally, thousands and thousands of hours have been devoted to the manual examination of large financial data sets (Madinger and Zalopany, 1999). Data mining technologies have the capacity to streamline this process. For example, the cluster analysis technique can be used as a data reduction method. Specifically, this technique is applicable when investigators are confronted with substantial amounts of data and there is a desire to group the data into a few categories based on similar attributes.

Neural network techniques can also reduce the time it takes investigators to generate case leads and identify suspicious activity patterns. These techniques can be used to drill down into large CTR and SAR data arrays to rapidly identify patterns similar to those the models were trained to find, and to discover new potential patterns as a result of model learning capabilities. The results from the analysis can then be analyzed by domain experts for further determination as to the quality of the leads and patterns identified.

The use of logistic regression could also save investigators time. This technique displays graphic views of the trends in the data, rather than the data itself. Furthermore, these trends are displayed in a multi-dimensional format using distinguishing shapes and colors. This can save investigators time as they would not be required to sort through the underlying data; the activity patterns would be already identified and graphically displayed. In other words, financial investigators will be presented with improved ways of viewing large masses of data in a manageable format. This makes pattern and trend identification more apparent and timely.

Second, it is feasible to suggest that data mining technologies could reduce problems in financial investigations that result from manpower shortages. Closely tied to the problems associated with time demands in money laundering investigations, manpower is always a major obstacle to successful lead and pattern identification. The use of data mining technologies could lessen the burden on already strained manpower resources. Because these methods have the ability to rapidly explore large financial data sets and identify case leads and money laundering patterns, it is conceivable that the inefficiencies in the current investigative processes could be reduced. Use of these technologies to rapidly and accurately identify leads and suspicious activity patterns could permit small well-trained units of domain experts to rigorously focus on analyzing outputs and leading investigations. This would be in stark contrast to the current use of under-trained and inexperienced investigators sifting through voluminous financial record sets attempting to isolate money laundering leads, patterns, and trends.

A third potential benefit of using data mining technology in money laundering investigations is the identification of more case leads, potentially more accurate case leads, and certainly more timely case leads. The use of a cluster analysis technique or a neural network technique offers the ability to identify previously undiscovered case leads. Domain experts have considerable experience working with financial data. They identify patterns and case leads to focus money laundering investigations. However, due to the sheer volume of financial transactions, it is humanly impossible to identify all leads and activity patterns given the stringent time constraints under which money laundering investigations proceed. The aforementioned methodologies can rapidly process data and expeditiously generate outputs for review by financial investigators.

The use of linear regression could provide financial investigators with more accurate and timely leads. This methodology is used to predict useful quantitative probabilities and delineate between strong and weak leads. For example, data from previously observed behaviors can be used to focus new investigative activities on the most promising locations at the most probabilistically promising day and time. This could remedy some of the problems in financial investigations that are associated with time and manpower shortages. In other words, if financial investigators can secure more accurate leads in a more timely fashion, they could more efficiently and effectively allocate resources.

LIMITATIONS ASSOCIATED WITH DATA MINING TECHNOLOGIES

While data mining solutions can provide a series of benefits to line level financial investigators, there are also a number of limitations associated with using technologically driven approaches in an investigative environment. First, investigative analysts must be trained to tune these customizable applications, analyze the results from the applications, and interpret the model results. For example, the performance of fuzzy logic methodologies and genetic algorithms will be limited if investigators or end-users do not commit additional prefield investigation time and computing resources to tune these custom AI models. These are additional activities by experts required to experiment with alternative model thresholds and 'what-if' scenarios electronically. Nonetheless, with their proper application, the expectation is that upfront investment of time required is more than recouped by more efficient and effective investigations while in the field.

The state of data mining technology today does permit software developers to create fairly intuitive applications that can present timely case leads and uncover suspicious money laundering activities and networks. Clearly, however, the utility of these techniques will be limited if financial investigators are not trained to understand (1) what the results from the various techniques mean, (2) what is necessary to improve the performance (generate better and more information) of each respective technique, and (3) which data mining approach will yield the results the investigator is looking for given varying case conditions. A second potential limitation to using data mining tools is directly related to the data used by these expert systems to generate case leads and identify suspicious activity patterns. The performance of these tools is limited by accessibility to financial data sets. Agency parochialism and unwillingness to share data has always been an enormous problem in the criminal justice system (Archambeault and Archambeault, 1984; Taylor, 1989; Bongiorno, 1996; Gondles, 2000; Lutz, 2000; Ratledge and Jacoby, 1989; Roberg *et al.*, 2000; Smith and Tolman, 2000). Law enforcement officials and financial investigators have experienced similar obstacles as they investigate money laundering operations locally, regionally, nationally, and internationally. The data sharing obstacle is not insurmountable, and the FATF is addressing this hurdle with notable success (Financial Action Task Force on Money Laundering, online). This progress will prove critical to financial

investigators who wish to use data mining tools to improve financial investigation processes. The inability to access financial records from multiple jurisdictions would clearly limit the effectiveness of data mining technology because these techniques require large data sets to train and refine the models. Moreover, access to partial data sets (i.e., information from sources within the USA only) will also restrict the robustness of any data mining technique.

Obtaining accurate and complete data sets is yet another obstacle that will restrict the productivity of data mining applications. Data integrity is an area that has consistently plagued financial investigations and it will also limit the performance of some data mining tools. For instance, the effectiveness of linear regression is primarily limited by the completeness, accuracy, and precision of the independent variables used in the regression model, as well as the consistency in suspects' behaviors. So, if an offender has multiple aliases and incomplete or inaccurate address information, the model may not identify a significant relationship between a particular offender and set of events. However, if sufficient information is available, this technique can identify a strong link between a suspect or suspects and certain behaviors. Here it is important for financial investigators to carefully evaluate the model results against their domain expertise to ensure that the best information is being identified and used to lead the investigation.

The utility of neural networks is also impacted by investigator accessibility to accurate financial data. If a sufficiently large data set with known patterns of money laundering activity is not available, and investigators are resolved to using relatively small incomplete data sets to train the model, the neural network will not be as robust in its identification of case leads and suspicious activity patterns. Fortunately, financial data sets usually consist of millions and millions of records, and are being made more accessible to financial investigators. Moreover, selection of the proper neural network architecture is important as this determines how the neural network is organized and thus capable of learning the desired relationships. Domain experts will prove critical for purposes of identifying the proper neural network architecture. They will need to examine model outputs against their experience to assist in the selection of the proper architecture.

Finally, one substantial obstacle that must be overcome on the part of law enforcement, and in particular financial investigators worldwide, is the unwillingness to change traditional ways of doing business (Taylor, 1989; Sparrow *et al.*, 1990; Skogan and Hartnett, 1997; Paoline, 2001). Police managers and administrators must address the organizational culture of law enforcement, move away from traditional reactive modes of service delivery, and embrace a proactive mindset. The use of data mining techniques and other information technologies can assist investigators solve crimes. These methodologies can also streamline police operations. However, without a dedicated effort to create an organizational environment that supports innovation, the use of data mining technology in investigation may be perceived by investigators as nothing more than a 'black magic' alternative to cracking cases.

TOWARDS AN INTEGRATED ANALYSIS ENVIRONMENT

The volume of CTR and SAR transaction records and the complexity of the relationships in financial data open the door to innovative techniques and methodologies that can aid financial investigators in generating timely, accurate leads. A variety of data mining techniques, especially neural networks and cluster analysis, are well suited. Use of such methodologies to track money laundering activities could drastically reduce the time and resources expended by financial

investigators and significantly increase the frequency with which financial data can generate valuable investigative leads.

A three-step research and development effort is appropriate for making the applicable AI techniques available to investigators. First, continued basic research will be fruitful to determine relevant relationships in financial transactions data. This will enhance the capabilities of the field investigators in identifying and subsequently investigating money laundering activities. In achieving this objective, data-driven AI tools, such as neural networks and inductive trees, are applied to discover relationships in the data, which can then be verified by subject matter experts. From this perspective, continued efforts should extend beyond the capabilities of existing expert system tools that exploit a priori expert knowledge derived in general. Moreover, since the data-driven technologies proposed can be made to operate in computationally efficient manners, this objective will guarantee a considerable reduction in the labor-intensive manual detection methods currently used by financial investigators worldwide.

Second, efforts should be focused on developing an inexpensive, Windows-based, automated software interface that can work with the new relationships discovered in the financial data to assist the field investigators in tracking down money laundering operations. Achieving this objective will put an inexpensive and effective tool in the hands of investigators that will significantly enhance their capabilities of identifying and further investigating illicit money laundering activities.

And third, an organized effort must be made to create pilot study programs through the widespread availability of secure, protected local area networks (LANs). This will allow the automated tool to be readily available to the appropriate federal, state, and local task forces. Moreover, this effort would be in line with and reinforce recommendations 30–40 of the FATF; specifically those aimed at strengthening international cooperation in the fight against money laundering (FATF online www.oecd.org//fatf/index.htm).

References

Abadinsky, H. 1997. *Organized crime* (6th ed.). Belmont, CA: Wadsworth.

Archambeault, W. and Archambeault, B. 1984. *Computers in criminal justice administration and management*. Cincinnati, OH: Anderson.

Baker, R. 1999. The biggest loophole in the free-market system. *Washington Quarterly*, **22**: 29–46.

Bentham, M. 1998. *Politics of drug control*. London: Macmillan Press.

Bongiorno, G. 1996. Information technology in policing. In R. Scherpenzeel (Ed.), *Computerization in the management of the criminal justice system* (pp. 295–303). New York: Criminal Justice Press.

Ciesielski, V. and Palstra, G. 1996. Using a hybrid neural/expert system for database mining in market survey data.

International Conference on Data Mining and Knowledge Discovery (Vol. 3, pp. 38–43).

Davis, A. and Goyal, S. 1993. Management of cellular fraud: Knowledge-based detection, classification and prevention. *Proceedings of the 13th International Conference on AI Expert Systems and Natural Language* (Vol. 2, pp. 155–164).

Eick, S. and Fyock, D. 1996. Visualizing corporate data. *AT&T Technical Journal*, **1**: 74–86.

Frayman, Y. and Wang, L. 1998. Data mining using dynamically constructed recurrent fuzzy neural networks. *2nd Pacific-Asia Conference PAKDD* (Vol. 1, pp. 122–131).

Genzman, L. 1997. Responding to organized crime: Laws and law enforcement. In H. Abadinsky (Ed.), *Organized crime* (p. 342). Belmont, CA: Wadsworth.

Gondles, J.A. 2000. Partnerships: A better way. *Corrections Today*, **62**: 6–7.

Haynes, A. 1996. *A comparative analysis of the effectiveness of three different approaches to combating money laundering: Policing central and eastern Europe*. Slovenia: College of Police and Security Studies.

Louie, J. and Kraay, T. 1999, March 9–10. Origami: A new data visualization tool. *Federal Data Mining Symposium and Exposition*, McLean, Virginia.

Lutz, W.E. 2000. Inter-local service agreements and trunked radio systems. *Sheriff*, **1**: 10–12.

Madinger, J. and Zalopany, S. 1999. *Money laundering: A guide for criminal investigators*. New York: CRC Press.

National Narcotics Intelligence Consumers Committee. 1996. The supply of illicit drugs to the United States. *Report 1996* [online]. Available: http://www.usdoj.gov/dea/pubs/intel/nnicc97.htm

Nossen, R. and Norvelle, J. 1996. *The detection, investigation, and prosecution of financial crimes*. Richmond, VA: Thoth Books.

Paoline, G. 2001. *Rethinking police culture: Officers' occupational attitudes*. New York: LFP Scholarly Publishing LLC.

Piatesky-Shapiro, G. 1996. An overview of issues in developing industrial data mining and knowledge discovery applications. *International Conference on Knowledge Discovery in Databases* (Vol. 1, pp. 89–95). Ratledge, M. and Jacoby, J. 1989. *Handbook on artificial intelligence and expert systems in law enforcement*.

Englewood Cliffs, NJ: Prentice-Hall. Refenes, A. (Ed.). 1995. *Neural networks in financial engineering* (Vol. 1, pp. 11–13). London: World Scientific.

Roberg, R.R., Crank, J. and Kuykendall, J. 2000. *Police and society* (2nd ed.). Los Angeles: Roxbury.

Robinson, J. 1996. *The laundrymen: Inside money laundering, the world's third largest business*. New York: Arcade.

Schmallenger, F. 1999. *An introductory text for the twenty-first century: Criminal justice today*. Englewood Cliffs, NJ: Prentice-Hall.

Scott, D. 1995, May. Money laundering and international efforts to fight it. *Public Policy for the Private Sector*, 48: 1–6.

Senator, T. 1995. The financial crimes enforcement network AI system (FAIS). *AI Magazine*, 4: 21–39.

Skogan, W. and Hartnett, S. 1997. *Community policing: Chicago style*. New York: Oxford University Press.

Smith, B. and Tolman, T. 2000. Can we talk: Public safety and the interoperability challenge. *National Institute of Justice Journal*, 243: 16–26.

Sparrow, M., Moore, M. and Kennedy, D. 1990. *Beyond 911: A new era of policing*. New York: Basic Books.

Standberg, K. 1997. Money laundering. *Law Enforcement Technology*, 4: 28–33.

Taylor, R. 1989. Managing police information. In D.J. Kenney (Ed.), *Police and policing: Contemporary issues* (pp. 257–270). New York: Praeger.

Trippi, R. and Turban, E. (Eds.). 1996. *Neural networks in finance and investing*. Irwin Professional Publishing.

Webster, B. and McCampbell, M. 1992. *International money laundering: Research and investigation join forces*.

Washington, DC: National Institute of Justice.

Williams, P. 2000. *Money laundering: Prospects for the future*. Office of International Criminal Justice [online]. Available: http://www.acsp.uic.edu/iasoc.crim_org/vol10_4/art_3m.htm

Yuanhui, Z., Yuchang, L. and Chunyi, S. 1997. Using neural network to extract knowledge from database. *1st European Symposium PKDD* (Vol. 1, pp. 377–383).

Zeldin, M. 1989. Detecting local business money laundering. *Asset Forfeiture Bulletin*, **2**: 1–6.