

# Connecting the Dots of Sensitive Terrorism Information for Homeland Security

<sup>1</sup>Ugochukwu Onwudebelu, <sup>2</sup>Jackson Akpojaro \*

<sup>1</sup>Department of Computer Science, Federal University of Ndufu, Abakaliki, Ebonyi State, Nigeria

<sup>2</sup>Department of Mathematics and Computer Science, Western Delta University, Delta State, Nigeria

anelectugocy@yahoo.com, \*jakpojaro@yahoo.com

## ABSTRACT

As society becomes more and more dependent on information and as criminals are increasing their cyber activities in their daily life, it becomes necessary to connect their dots together to track them in this information age. Terrorism is not confined to one country and it has no borders or boundaries. The escalating magnitude of this threat is evident from the increasing rate of terrorist attacks against innocent people, especially in the Northern part of Nigeria. As we are seeing, one of the major concerns of many nations today is to identify and foil terrorist attacks emanating from different angles. Consequently, data mining which is being used for almost everything from improving service or performance to detecting specific identifiable terrorist threats is employed. Defeating terrorism requires quick intelligence machinery that operates more effectively and makes use of advanced information technology such as data mining and automated data-analysis techniques for a successful fight against terrorist as well as collaboration in data-sharing program between the three levels of government: federal, state and local. In this paper, we are looking at the need to design support information sharing among these levels of government. So that the government, as a whole will use its power to affect the lives of individuals increasingly with regards to safeguarding lives and properties.

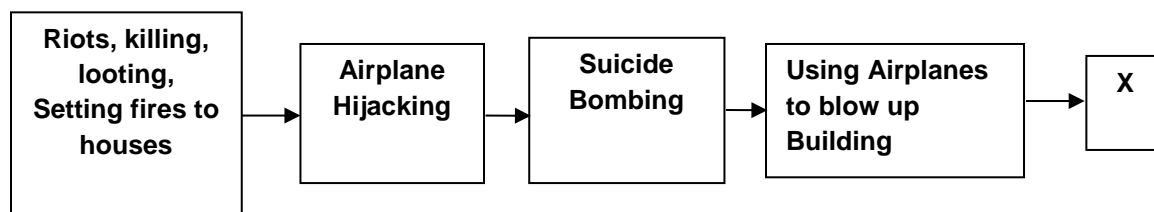
**Keywords:** data mining, homeland security, threats, profiling, data set

## 1 INTRODUCTION

People have always depended upon Information Technology (IT) of some type, beginning with smoke signals in ancient days and turning into network-based computer systems today. Nowadays, the computers control power, oil and gas delivery, communications, transportation, banking and financial services. Furthermore, they are used to store and exchange vital information, from publicly known facts to well kept secrets [5]. It is clear now that the threat we face from terrorism is far different from Cold War threats and requires adjustments to our

approach to national security threats, to intelligence collection and analysis. Unlike Cold War adversaries, the terrorists are loosely organized in a diffuse and nonhierarchical structure [2]. There is terrorism everywhere and carried out by people from different countries, speaking different languages. Thus, they are everywhere, in every country and without regards to human life and property. The terrorists activities are of different natures (see Figure 1). X in the figure represents the unknown nature of the next-generation attack.

Data mining has been defined as the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [7]. It is the science of extracting useful information from large data sets or databases. Data mining is emerging as one of the key features of many homeland security initiatives. It is often used as a means for improving program performance, detecting fraud or abuse (waste), assessing risk, product retailing (to reduce costs), analyzing scientific and research information, managing human resources, detecting criminal or terrorist activities or patterns, and analyzing intelligence in both the private and public sectors.



**Figure 1. Sophistications and progressive nature of terrorism**

Consequently, the government is relying increasingly upon data mining programs, namely the use of computing technology to examine large amounts of data to reveal patterns and identify potential wrongdoing [9]. Detecting combinations of low-level activities—such as illegal immigration, money transfers, use of drop boxes and hotel addresses for commercial activities and having multiple identities—could help predict terrorist plots [2]. Used properly, data mining can provide a valuable tool for the government to uncover fraud or criminal activity.

While all traditional intelligence collection methods remain critically important, understanding the terrorists and predicting their actions requires us to rely more on making sense of many small pieces of information. Given the ethnic and religious makeup of the 9/11 perpetrators and other recent terrorists such as the 2009 Christmas Day bombing attempt by Farouk Abdulmuttallab, various bombing campaign against Nigeria military and schools in the Northern part of the country by the dreaded Boko Haram sect whose leader Sheikh Abubakar Shekau has claimed responsibility etc., based on these, the program might flag Muslim men. However, would-be terrorists can come from any racial or religious groups or countries of origin, and thus such racial profiling would not only unfairly target certain individuals. Although other variables that may be included as part of a data mining algorithm [9] such as a passenger travelling on a one-way ticket or carrying a large quantity of cash—may similarly generate under- or over-inclusive lists, we must be especially careful in the case of racial, ethnic, and religious

classifications. This data analysis tool is very important in the war against terrorism, by using government watch list information, airline reservation records, and aggregated public record data, link analysis could have identified all 19 September 11 terrorists - for follow-up investigation - before September 11 [2]. In the wake of 9/11, governments around the world have developed tools useful in mining data. Furthermore, governments have built or are building thousands of databases and are deploying hundreds of data mining applications to law enforcement agencies, communications and intelligence data for terrorist, therefore Nigeria cannot lag behind in the fight against national security threats.

### **1.1 Homeland Security**

Homeland security is very essential at the moment to secure a nation from the many terrorist threats facing it, both domestic and international terrorists. Nigeria government must put on concerted effort to prevent terrorist attacks within our three-tiers of government as well as reduced the vulnerability to terrorism. Consequently, officials involved in homeland security may take into account specific, credible information about the descriptive characteristics of persons who are affiliated with identified organizations that are actively engaged in threatening the national security. Information from the “watch list” must be distributed throughout the government, including police, department of military intelligence (DMI), immigration, customs, consular offices overseas, state and local law enforcement agencies [8] for prompt action against terrorist activities. For effective homeland security, the federal government needs to support the development of state and local information fusion centers.

## **2 TYPE OF DATA MINING**

Data mining is the broad term used to refer to many types of activities involving data processing. Data mining is divided into two: Pattern-Based Data Mining (PBDM) and Subject-Based Data Mining (SBDM). In PBDM, Such pattern-based systems learn over time by examining the data, comparing the data to a model, and then searching databases for patterns matching the revised model [6]. Federal money-laundering investigators, for example, might input information about financial crimes and criminals into a sophisticated data mining system, which would review banking data for transactions or accounts that share suspicious attributes with the criminal data points. While in SBDM, the data are simply scanned for items or events meeting specified parameters in “subject-based” queries [3]. For example, anti-graft officers might start with a known suspect and use a multi-jurisdictional database to search for information about that suspect, such known associates. A major goal in research on data mining for counterterrorism, for example, is not only to identify terrorist “signatures,” but also to find ways to separate those patterns of activity from all other “noise” in databases [2]. Although, these distinctions are often blurry, however many data mining systems use both subject-based and pattern-based techniques in its operations. The differences between PBDM and SBDM are illustrated in Table 1.

**Table 1: The Differences Pattern-Based and Subject-Based Systems**

<b>Type of Data-Analysis Technique</b>	<b>Pattern-based Data Mining</b>	<b>Subject-based Data Mining</b>
Definition	It is the use of “pattern-based” searching to uncover novel patterns or relationships in large sources of data.	It is the use of “subject-based” searching to simply scanned for items or events meeting specified parameters.
Type-Based Query	Pattern-based queries involve identifying some predictive model or pattern of behavior and searching for that pattern in data sets.	Subject-based queries start with a specific and known subject and search for more information.
Identity	It can help reveal patterns and relationships. However, it does not tell the user the value or significance of these patterns. The user need to interpret the output that is created.	The subject could be an identity such as a suspect, an airline passenger, or a name on a watch list, a place or a telephone number etc.
Example of software	Tableau Software	Non Obvious Relationship Awareness (NORA™) software
Uses	To detect money-laundering activity, to detect credit card fraud	To prevent fraud, cheating, and theft from casinos.
Policy Difficulties	More, because Pattern-based queries are less familiar in the law enforcement and intelligence worlds in that they do not arise from a particular interest in a person, place, or thing.	Fewer because they are more like the kinds of inquiries that are common in intelligence and law enforcement practice
Usefulness in counterterrorism	It has potential for counterterrorism in the longer term, if research on uses of those techniques continues.	More effective in the counterterrorism realm
Link Requirement	It searches do not require a link to a known suspicious subject.	It searches do require a link to a known suspicious subject.

### 3 METHODOLOY

The information on terrorist threats we have presented here has been obtained entirely from unclassified newspaper articles, online documents, conference papers, journals as well as news reports. Our focus is to illustrate how data mining could help towards combating terrorism by reason of strong information sharing at the three-tiers of government (which involves connecting the dots) especially in Nigeria where the terrorist group Boko Haram is claiming more and more lives weekly. In the context of homeland security, data mining is often viewed as a potential means to identify terrorist activities, such as money transfers and communications, and to identify and track individual terrorists themselves, such as through travel and immigration records. The data of interest to the suspect or person of interest include names, addresses, phone numbers, date of birth, height, weight, and social security numbers (in countries that make use of it) drawn from various sources. Others may include race or religion, although it is sometimes not encouraged because of the discriminatory effect of racial

classifications. Consequently, racial profiling reduces individuals' trust in the government. Innocent individuals who are marginalized due to racial profiling may be far less likely to participate in public affairs, or to cooperate with the government to combat threats to national security in the future.

### **3.1 Events and Entities Required in Database**

An attempt to find interesting events from the database i.e. events that require further action on their part, such as checking suspicious character. Unfortunately, the lack of a concrete database whose data emanate from the local to the state can cause important events to be buried within some millions events. In such a centralized fusion center (see Figure 2), a search can be made, and the data investigated to see how closely linked data are to an individual's identity. To avoid unnecessary burdens on the government, notice should be undertaken only when an individual has been subject to a specific action or classification and it is feasible to locate and trace the individuals or better still to monitor the suspect. Accurate identification at each level not only is critical for determining whether a person is of interest for a terrorism-related investigation, it also makes the government better at determining when someone is not of interest, thereby reducing the chance that the government will inconvenience that person.

We need to start gathering information about various people including those who seem most innocent but may have ulterior motives. What data should we collect? The individual records may include the following data: names, addresses, phone numbers, date of birth, height, weight, Postal Service address, hotel addresses, driver's license, driver's license pictures, professional licenses, names of neighbors and relatives, motor vehicle information, telephone number (contact number), social security numbers and criminal records. If we know that someone has had a criminal record, then we need to be more vigilant about that person. In summary, we need to group the individuals depending on say where they come from (nationality, state, local), what they are doing, who their relatives and associates are etc. This information could include information about their behavior, their travel records, where they have lived, their religion and ethnic origin, etc.

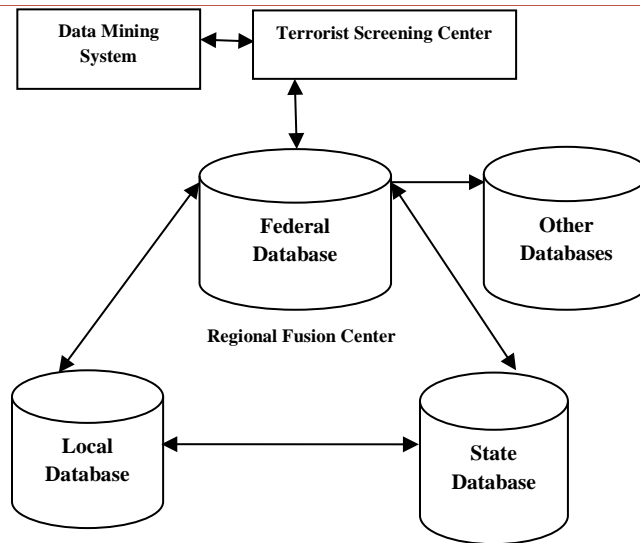


Figure 2: Information sharing in the Three-Tiers of Government

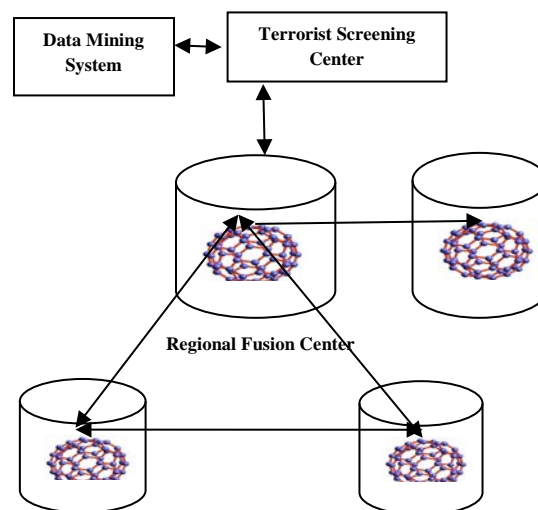


Figure 3: Connecting the same Dot in different Databases

Some people may have more suspicious backgrounds than others. The more information that is available, the more accurate the identity resolution process becomes and the easier to connect the dots. By omitting some of this crucial information we may not have the complete picture. This information amongst other things should include motor vehicle registration, dependents' information, passport information, field of study, and employment information, corrections information, credit card information, demographic information, meal information (which can hint to religion), sexual offender information, record on credit-card purchases, plane flights, e-mails, websites, housing, home and business addresses. From a technology point of view, we need complete data not only about individuals but also about various events and entities.

### 3.2 Inter-relation Ship

An increasing amount of such data mining is occurring at “fusion centers,” centers within each state that bring together federal, state, and local law enforcement personnel to share information and coordinate activities (see Figure 4). Through these fusion centers, the federal government has acquired data from state and local law enforcement databases to improve information sharing and availability among law enforcement and intelligence agencies. While more efficient sharing of data can undoubtedly aid law enforcement efforts, the unlimited scope, lack of transparency, and lack of oversight for the program create significant risks to civil liberties (see section on Privacy). A proper inter-connection must be designed and made to focus exclusively on identifying and preventing terrorism threats.

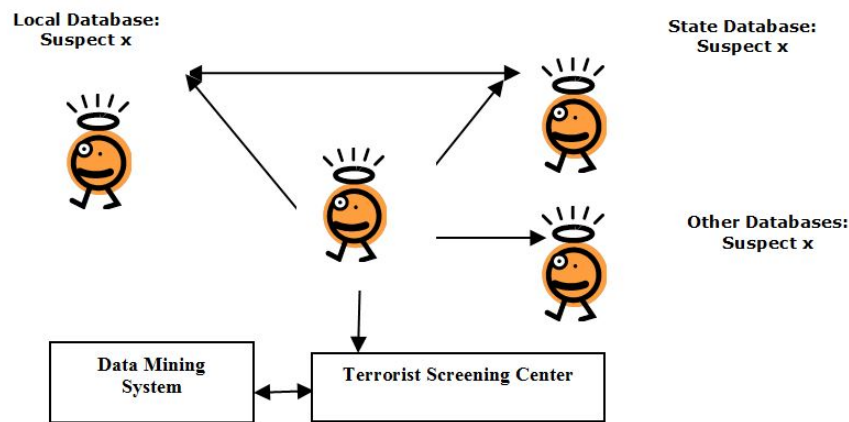


Figure 4. Connection Dots Network

Consequently, among the insights from this research it is more productive and less prone to error to follow connections from known starting points – subject-based (see Table 1). The system was designed to use information acquired from the three-tiers of government sources and other agencies such as network providers to do the proper connections and thus looks for anomalies or patterns that indicate certain behavior peculiar for terrorist or criminal threat.

### 3.3 Database Connections @ the Three-Tiers

One initial potential benefit of the data-analysis process is that the use of large databases containing identifying information assists in the important task of accurate identification [8]. When information gathering originate from the local to the state and to the federal, then such information makes it far easier to resolve whether two or more records represent the same or different people (identity resolution). The Nigerian government needs a National Identification Number (NIN) to make profiling more accurate. Therefore, we need to introduce NIN into our system, issuing it to every child born in the country as well as to every Nigerian, including those citizens abroad. The NIN will help in identity resolution at determining when a person in question is not the one suspected of terrorist acts, thereby potentially reducing inconvenience.

In tracing the data, the dots of the terrorist might be at any level of the three-tiers of government. Consequently, one has to connect all the dots (see Figure 3). Essentially one builds a graph structure based on the information he or she has from the three levels. From these dots, we need to find out who these people are by analyzing their connections and then develop counter-terrorism solutions. Note that counter-terrorism is mainly about developing counter-measures to threats occurring from the terrorist's dot activities. Even the Banks search databases of credit card transactions can make the connection from the dots, some of which are known to be fraudulent, and determine, through data mining or otherwise, the patterns of fraudulent activity. Indeed, in the complex world of counterterrorism, application of data-mining and the right connection models and related techniques are likely to be useful at several stages of a multistage process of developing a complete picture out of many "dots." Based on a successful connection to the dots, government actors may want to take action based on the results and other results from data-analysis queries. This action could include detention, arrest, or denial of a benefit.

It should be noted at this point that although data mining could contribute towards counter-terrorism. Nevertheless, we are not saying that data mining will solve all our national security problems and threats. However the ability to extract hidden patterns and trends from large data sets is very important for detecting and preventing terrorist attacks and in homeland security. The connecting dots can be derived from the suspect travel patterns or eating patterns or buying patterns or behavior patterns. We must be able to find little dots of data in a sea of information and make a picture out of them. By omitting some information we may not have the complete picture as stated above. But from the breadth of access to information and quality analysis, a number of clues, if recognized, combined, and analyzed might have given us enough to track down the terrorists and stop their plan. The data mining as well as the databases at the three levels must emanate from the same agency for easy interoperability and implementation issues.

### **3.4 Terrorist Identities Classification**

Most observers believe that data mining can improve government performance if used appropriately. Data-mining and automated data-analysis techniques are not a complete solution [2], but they are powerful tools that help government in terrorism prevention. They can assist investigators in matching crime scene evidence to other crimes or suspects or finding known associates or other information about persons of interest. The federal government should have a Terrorist Identities Database (TID) which should be related to the identities of individuals known or appropriately suspected to be or have been involved in activities constituting, in preparation for, in aid of, or related to terrorism, including those that in their "watch lists" and "No Fly" list. A non-exclusive list of types of conduct that will warrant entry into TID includes persons who:



- Commit domestic or international terrorist activity;
- Prepare or plan domestic or international terrorist activity;
- Gather information on potential targets for domestic or international terrorist activity;
- Solicit funds or other things of value for domestic or international terrorist activity or a terrorist organization;
- Solicit membership in a domestic or international terrorist organization;
- Provide material support, i.e. safe house, transportation, communications, funds, transfer of funds or other material financial benefit, false documentation or identification, weapons, explosives, or training;
- Are members of or represent a domestic or a foreign terrorist organization.

Terrorist Screening Center, which is a critical instrument for homeland security, supports screening processes to detect and interdict known and suspected terrorists at home and abroad – with the information stored in the local, state and federal databases the terrorist will be apprehended. The federal government needs to develop tools that will help to mine data from local, state and if necessary from the private sector such as Internet service provider, Telecommunications Companies, etc. since they have become a repository of a host of personal data. This would be the ideal solution and the research challenge is to develop such a data miner. We are recommending the federal government to create a Homeland Defense Agency (HDA) that will connect the dots at the three-tiers of government so as to have access to the three databases, which will create graphs and make the links that will connect to the dots of sensitive information to a potential and identifiable terrorist. With access to the HDA database, law enforcement investigators can look up vast amounts of personal information culled from government and if need be from commercial databases. It is paramount to ensure that consistent, accurate and complete terrorist information is disseminated to frontline screening agents in a timely manner. Not only will a “flagging” result in greater surveillance, but it could also result in detention, interrogation, or otherwise intrusive investigation. For a person who is innocent, these events will be negatively life-altering.

#### **4 NATIONAL BOUNDARIES ISSUES**

Today it is Nigeria and tomorrow it may be another country. Due to the pervasiveness of technology and Internet connectivity, the scope of terrorism as well as cybercrime incidents are often perpetuated across national boundaries. Safeguarding the borders is critical for the security of our nation from International terrorist. According to the Federal Bureau of Investigation (FBI), international terrorists include those persons who carry out terrorist activities under foreign direction. Attacks on borders as well as transportation are increasing in alarming rate recently. Thus, there are so many discussions that are related to securing the borders and transportation industry. There are threats at borders from illegal immigration, kidnapping, prostitution, child pornography to gun and drug trafficking as well as human

trafficking to terrorists entering a country, attacking and taking refuge at the mountains and thick forests of neighboring countries.

We are not saying that illegal immigrants are dangerous or are terrorists. Nonetheless, they have entered a country without the appropriate documents and that could be a major cause for concern. As for drug trafficking at the borders [10], drug can cripple a nation, corrupt its children, cause havoc in families, damage the education system and cause extensive damage to the brain and economic mainstay of a country. Consequently, Nigeria Immigration Service (NIS) and Nigeria Customs Service (NCS) have to collaborate with Chad, Benin, Niger, and Cameroon in joint border patrol as part of the efforts to winning the war against terrorism in Nigeria. Nigeria needs to seek the mandate of the respective Governments to establish joint patrol teams along the common borders to promote security [1]. The insurgent are attacking Nigerian citizens and running away to these neighboring countries. Furthermore, we have to protect our borders so that there are no additional problems to our nation.

#### **4.1 False Identification Issues**

Note that the terrorists like other criminals can often anticipate the factors that law enforcement will use to profile them and will circumvent them quickly enough. The potential for false identification cannot be neglected as terrorist may attempt to deliberately modify their methods to avoid mimicking past terrorist plots undermining pattern-based data mining methodologies. A false positive as defined in [2] is when a process incorrectly reports that it has found what it is looking for while a false negative is when it incorrectly reports that it has not found what it is looking for. Thus, false positives and false negatives are inevitable and they both increase costs to the government and create public skepticism about the value of security measures. Nevertheless, it should be noted that it is very vital to make sure that the data mining tools produce accurate and useful results. For example, if there are false positives, the effects could be disastrous for various individuals while false negatives could increase terrorist activities. Even if the government later corrects its mistake, the damage to reputation could already be done, with longer term negative consequences for the individual.

#### **4.2 Privacy Issues**

Increased government access to and use of information brings significant benefits, but also increases the risk of encroachment on constitutional rights and values—including privacy, freedom of expression, due process, and equal protection [9]. The public is always pessimistic concerning the idea of anyone knowing too much about their personal lives as well as their electronic life. Because of this use of personal information, the business world and the government are working hard to find a way to mine data without interfering with legal, privacy, and security concerns that are raised by the public. This has resulted in the coming together of individuals from different professions such as counter-terrorism experts, civil liberties unions and human rights lawyers to find a solution to the issue of infringing on individual's privacy.

That is, gathering information about people, mining information about people, conduction surveillance activities and examining e-mail messages and phone conversations without due processes. But how can we combat terrorism effectively without trampling on the privacy of individuals? What is more important? Protecting the nation from terrorist attacks or protecting the privacy of individuals? This is one of the major challenges faced by counter-terrorism experts, civil liberties unions and human rights lawyers. The same questions were asked in [10]. That is, how can we have privacy but at the same time ensure the safety of our nations? What should we be sacrificing and to what extent? The challenge is to provide solutions to enhance national security but at the same time ensuring that the privacy of individuals are not compromised. However, technology is increasingly blurring the lines between spheres in which people commonly do or do not expect privacy. For instance, individuals have no choice but to disclose information to a third party in order to be able to participate in basic aspects of modern society, such as online banking, storing electronic business or financial records online, communicating by phone or email, or using a credit card to make purchases [9]. Federal agencies should collaborate to adopt government-wide, written, defined standards for the acquisition, sharing, and use of data so that the data at all tiers of government is protected. Operators who do not follow the standards, or who otherwise misuse or abuse personal data or data mining systems, should be subject to civil or criminal penalties.

## 5 RECOMMENDATIONS

- Federal Government should explore the beneficial framework involving the partnership between state and local governments in information sharing of databases.
- There is a need for redress mechanisms for those aggrieved by a data mining activity, that is, government should establish a system of appeal and redress for individuals misclassified or harmed.
- Duplicate records, incomplete records, timeliness of updates, and human error all create data integrity problems. As a result, qualified and trained database administrator must be stationed to handle the database of all three-tiers since the outcome of data mining can only be as good as the underlying data.
- Where feasible, especially at the local and state levels (grass-root), individuals should have the opportunity to review their information held by government but should not be permitted to update the information. This will provide a sound means of ensuring that the data in the database are accurate, reliable, timely, and complete. It will also preempt potential harm that may result from the use of inaccurate or unreliable data. Any errors noted by individuals relating to their personal data should be promptly reported and corrected by the appropriate agency and the database synchronized on regular basis to provide dynamic and timely information.

## 6 CONCLUSION

As almost everyone now recognizes, the fight against terrorism requires the government to find new approaches to intelligence gathering and analysis. Data mining tools are effective and powerful techniques in the war against terrorism especially in the complex world of counterterrorism where conclusions and decisions must be made to stop the potential harm of catastrophic terrorism as well as detecting suspected terrorists. However, it is a mistake to view data mining and other linked tools such as automated data analysis as complete solutions to security problems and threats. Their strength is as tools to assist analysts and investigators at the three-tiers of government. Furthermore, data-mining and automated data-analysis techniques can find links, patterns, and anomalies in large data sets that humans could never detect without this assistance, which help investigators of terrorism form the basis for further human inquiry and analysis.

In this paper we have demonstrated how the local, state and federal governments can collaborate and support sharing timely intelligence information. If our recommendations are carried out properly, it will help the government to impact our lives and rights of individuals increasingly and serve our essential national values. Although private data mining is beyond the scope of this paper, it still implicates similar privacy concerns and therefore we recommend that federal, state, and local governments contemplate private-industry regulation to protect individual liberty interests. Thus, it has been shown that data mining can contribute towards the battle against terrorism, further enhance defense mechanisms of a nation and can advance counterterrorism goals. Apart from combating terrorism, for actions or classifications that are made as a result of data mining, such can help for instance flagged individuals during auditing for tax or any other fraud. In USA, for example, a data mining program has helped uncover millions of dollars in Medicare fraud, combating fraud and auditing for compliance [4].

We are not saying that data mining solves all the problems associated with terrorism rather it has the capability to extract patterns and trends, often previously unknown, we should certainly explore the various data and web data mining technologies for counter-terrorism. Finally, privacy rights can be implicated by inappropriate sharing and downstream uses of information gleaned from data mining. As a result government should incorporate technical and administrative measures to limit access to or availability of personal data when it has to do with non-terrorism-related investigations. This will help checkmate government employees who can abuse database access and look for information on the famous or infamous.

## REFERENCES

- [1]. Alohan, J., Terrorism: Nigeria, Chad, Benin , Cameroon, Niger In Joint Border Patrol Deal, Page 4, Thursday, march 27, 2014. <http://www.leadership.ng> , No. 2160.

- [2]. DeRosa, M., Data Mining and Data Analysis for Counterterrorism, Center for Strategic and International Studies (CSIS) Press, 2004.
- [3]. DHS Privacy Office, Data Mining: Technology and Policy: 2008 Report to Congress, pp. 31-32
- [4]. Department of Homeland Security, Privacy Policy Guidance Memorandum, No. 2008-01, Dec. 29, 2008, available at [http://www.dhs.gov/xlibrary/assets/privacy/privacy\\_policyguide\\_2008-01.pdf](http://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf) (memorializing DHS adoption of the FIPPs).
- [5]. Kumar, V., Lazarevic, A., and Srivastava, J., Workshop on Data Mining for Cyber Threat Analysis, IEEE International Conference on Data Mining, Maebashi TERRSA, Maebashi City, Japan, 2002.
- [6]. Minow, N. N. and Cate, F. H., Government Data Mining, at 4, <http://ssrn.com/abstract=1156989>, in McGraw Handbook of Homeland Security (2008); National Research Council, Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment, at 22 (National Academies Press), 2008.
- [7]. Seifert, J. W., Data Mining: An Overview, CRS Report for Congress Received through the CRS Web, December 16, 2004.
- [8]. Shenon, P., Inspection Notes Errors in Terror List, 2007.
- [9]. Sloan, V. E., and Sharon B. F., Principles for Government Data Mining Preserving Civil Liberties in the Information Age. 2010.
- [10]. Thuraisingham, B., Data Mining for Counter-Terrorism, The MITRE Corporation Burlington Road, Bedford, MA, USA., 2011.