COMMERCIALIZATION OFFICE - PILOT OPERATIONAL REQUIREMENTS DOCUMENT

Predictive Modeling for Counter-Improvised Explosive Devices

February 2009

Point of Contact:
Mark Protacio
Commercialization Office
SETA Support
Mark.Protacio@associates.dhs.gov

Contents

1. General Description of Operational Capability

1.1. Capability Gap

This operational requirements document (ORD) addresses the capability to predict the threat of an IED attack, identified by the Counter-IED Capstone IPT. It also covers a number of technology needs identified to further data fusion from law enforcement, intelligence partners and other sources to support the common operating picture.

1.2. Overall Mission Area Description

The Department of Homeland Security (DHS) plays a major role in fulfilling Presidential Directive/HSPD-19 (Combating Terrorist Use of Explosives in the United States) including national policies, strategies and implementation plans for the prevention and detection of, protection against and response to terrorist use of explosives in the United States.

Terrorists have repeatedly shown their willingness and ability to use explosives as weapons worldwide and there is ample intelligence to support the conclusion that they will continue to use such devices to inflict harm. The threat of explosive attacks in the United States is of great concern considering terrorists' ability to make, obtain, and use explosives, the ready availability of components used in IED construction, the relative technological ease with which an IED can be fashioned and the nature of our free society.

It is the policy of the United States Government to counter the threat of explosive attacks aggressively by coordinating Federal, state, local, territorial, and tribal government efforts and collaborating with the owners and operators of critical infrastructure and key resources to deter, prevent, detect, protect against and respond to explosive attacks, including the following:

(a) Apply techniques of psychological and behavioral sciences, such as social network theory, in the analysis of potential threats of explosive attack;

(b) Use the most effective technologies, capabilities, and explosives search procedures and applications to detect, locate and render safe explosives before they detonate or function as part of an explosive attack, including detection of explosive materials and precursor chemicals used to make improvised explosive or incendiary mixtures;

(c) Apply all appropriate resources to pre-blast or pre-functioning search and render-safe procedures, and to post-blast or post-functioning investigatory and search activities, in order to detect secondary and tertiary explosives and for the purposes of attribution;

(d) Employ effective capabilities, technologies and methodologies, including blast mitigation techniques, to mitigate or neutralize the physical effects of an explosive attack on human life, critical infrastructure, and key resources; and

(e) Clarify specific roles and responsibilities of agencies and heads of agencies through all phases of incident management from prevention and protection through response and recovery.

1.3. Description of the Proposed Product or System

The proposed solution shall employ the knowledge and understanding gained in the military environments such as Iraq and Afghanistan to model and take action against IED network activities in the United States. It shall enable investigators to disrupt networks by expanding analysis and investigation beyond the groups and individuals that place devices to analyze and target the finances, materiel and supply line of parts, and “the brains” that build and deploy IEDs.

DHS knows that the insurgents who seek to place IEDs in the United States (as they do in other parts of the world) are often supported by organized networks that finance their operations, supply critical elements for the production of IEDs, create the devices and plan and execute attacks. The proposed solution shall implement powerful analytics to gain critical, data-driven insight into the structure, character, interactions and methods associated with those networks. By analyzing data from a myriad of sources, the new solution shall identify and analyze the linkages between individuals and groups that may indicate a support network.

1.4. Supporting Analysis

DHS has undertaken an array of activities designed to prevent the detonation of IED/VB-IED/suicide bombs inside the United States and against American interests abroad. The department is aggressively working to focus on identifying and attacking the threat before terrorists have the capability to detonate a device. That begins with attacking the foundation of the threat—the social, operational and financial networks. Critical to the efforts is developing an integrated, cross-agency data-driven foundation of intelligence as the basis for deterring and incapacitating those who supply/obtain the funds for IEDs, identifying the organization planning to manufacture and plant the IED and intercepting the gathering and procurement of materials for the IED.

1.5. Mission the Proposed System Will Accomplish

The proposed solutions is envisioned to be a seamless, transparent and an integrated combination of COTS software, training and services that form an intelligence collection and analysis system that will help uncover and target the operational, financial and social networks involved in IED deployment in the United States. The solution shall address the challenges of data access, integration, quality and management of data coming from multiple government agencies and publicly available sources. In the modern and developed world, where most of the explosives/IED support networks operate, government agencies and the private sector generate unprecedented volumes of data. Customer profiles, organizational operational performance and personal behavior of individuals are monitored by multiple service providers. Some data resides in structured form in databases or exists as real-time streams. Some exists in unstructured form, for example as e-mails, electronic documents or media files. Whatever the form, there exists potential to transform these data into relevant intelligence to improve investigative decision-making. A proposed solution shall integrate existing data from all relevant sources, and with its advanced analytics and reporting capabilities, provide actionable information to U.S. investigators in the full range of Federal, state, local and tribal jurisdictions.

This must be a cross-agency solution that is designed to deliver a broad range of intelligence products within a multi-level environment (Federal, state, local and tribal jurisdictions) to provide the full community of decision makers, analysts and investigators with better information to address potential threats.

Because first responders are an integral part of a tactical, pre-initiation response to an immediate threat, the proposed solution must address the appropriate type and level of information that would support those contingencies and how that information would be shared at the first responder level.

A proposed solution can use an integrated suite of tools, including but not limited to data integration and management, data and text analysis, predictive modeling and optimization and social network analysis coupled with link analysis. Analysts and other end users will receive detailed intelligence developed using data driven investigative techniques and link analysis based on social network theory. Analysts will be provided with client tools, such as customizable report creation and delivery capabilities that provide intelligence in the most appropriate format for decision makers and other users. The solution can be customizable, if desired, at all levels of security classification. The data will be searchable via a graphic user interface that will allow the investigator/analyst team to search for unusual behaviors and complex sequences of behaviors across records.

1.6. Operational and Support Concept

1.6.1. Concept of Operations

A solution will bring together key elements of intelligence needed to enable analysts, investigators and decision makers to make data driven decisions to more effectively perform their mission.

Specifically, the proposed solution shall:

- Link disparate, cross-agency data sources and integrate required elements

- Enhance data quality and accuracy

- Develop information across large volumes of integrated data

- Use data/text mining, predictive modeling and other advanced analytics methodology to provide insight to relevant data

- Expediently operationalize intelligence

- Identify suspicious social, financial and operational networks that may be appropriate for further analysis or investigation

- Communicate actionable information out to decision makers and investigators via the Internet, LAN/WAN, email, etc.

- Enable the agency to better detect and defeat potentially dangerous networks

A solution must integrate the efforts of analysts and decision makers at a cross-department/cross-agency level in a “fusion center” type environment. Sources of record will be made available so that the solution can automatically pull data on a near real time basis to update the intelligence available. A proposed solution will work within the specified DHS IT environment and will pull data from existing systems. Access to data sources will be granted by the responsible agency.

1.6.2. Support Concept

The responsible department/agency will consider options for the implementation and sustainability of the proposed solution.

The data integration and management, along with the analysis and modeling, social network development, linking and scoring functions shall be maintained at the agency level, with analysis and reports made available over the appropriate networks to users based on role or persona. Analysts, based on their role and mission requirements, will be given additional analytical capabilities to better perform their responsibilities.

A proposed solution will support the full range of services required for sustainability of the system. This shall include data integration and cleansing, data and text mining, predictive modeling, social network development linking and scoring. A solution will include proposed actions that the agency could take to develop the appropriate skills within the organization, particularly those related to data extraction, cleansing, integration and intelligent storage.

Training in the operation of the system, both for the initial implementation and for long term sustainability, shall be provided as part of the solution, as well as tailored courses, delivered on site, that focus on specific agency issues and requirements.

2. Threat

The potential threat to the United States from improvised explosive devices, vehicle-borne IED (car bombs) and suicide bombers is well established. Incidents since the beginning of 2000 in Bali, Madrid, London, Libya, as well as the attacks on the U.S. in Oklahoma City on 19 April 1995 and the World Trade Center bombing on 26 February 1993 attest to the potential threat and difficulty in protecting against them. This is not a new phenomena, as seen in the actions of the Red Brigade and Bader-Meinhof Gang directed against U.S. interests in Europe in the mid 1970, the attack on the U.S. Marine Barracks in Beirut in 1983 and others. Current intelligence and law enforcement estimates predict that these types of attacks against the population of the United States are inevitable.

Specific types of attacks could be the type of IED/VB-IED/suicide bombing widely employed in current combat zones and around the world, but could also include use of explosive devices combined with commercially available, stolen or smuggled biological, chemical or radiological agents to cause further loss of life, widespread panic and economic damage.

Nearly every incident of IED/VB-IED/suicide bombing involves groups or networks of individuals acting in concert. While these networks may vary in type and complexity, they have common characteristics, such as the need to communicate, fund operations, procure materiel and travel to accomplish their objectives. These activities leave transactional records in databases legally maintained by government and commercial entities. By leveraging and extending insight into these data sources, it is possible to assemble a threat profile to protect against future attacks.

3. Existing System Shortfalls

Current systems fall short in the following areas:

a) Lack of the capability to integrate data from disparate sources. Data is contained in multiple systems within disparate organizations. It is often on different platforms and in different environments with different security and access requirements. Some is in transactional systems such as Oracle, SAP, DB2, Microsoft desktop applications and others, some in proprietary databases, some in legacy systems built in FORTRAN, COBOL or ADA. Integrating the sources is a complex technical problem, complicated by the various internal departmental/agency policy and cultural issues.

b) Volume of data to be analyzed. Not only is data in various agencies, formats, platforms and environments, the amount of data that should be considered in a comprehensive program is quite substantial, with gigabytes if not terabytes available for analysis. Since many of the transactional sources are updated in real time and others on a daily basis, the volume of raw data to be extracted, integrated and analyzed as required to provide timely, actionable information to analysts and investigators is a significant challenge.

c) Solutions lack scalability and robustness. Current less-flexible and less-capable systems have issues such as how new agencies or data sources can be incorporated into the data integration regime. Because current solutions are typically single agency efforts, the requirements for scalability and robustness are typically not addressed.

d) Lack of advanced analytics. Agencies typically do not have the capability to apply high end or advanced analytics, such as data indexing and profiling, data and text mining, predictive modeling, forecasting and optimization to their mission requirements. The types of network analysis and network linking required cannot be accomplished using multiple purpose, less sophisticated technology.

e) Absence of a foundation in social network theory. Department/agency personnel do not have a thorough grounding in the theory of how social networks interact and change patterns of behavior. This hampers their ability to gain maximum intelligence from existing data. The use of social network theory domain experts, thought leaders, academicians, investigators and analysts, operating within the appropriate technology environment is not optimal.

4. Capabilities Required

4.1. Operational Performance Parameters

The performance metrics included as part of the Threshold (T) and Objective (O) Values are based on 10 government data sources, 300 total users – 50 of them concurrent – located in 3 locations within the United States.

		Objective Value	Threshold Value
1		Data Integration	
	a	Integrate, cleanse and store data from multiple sources.	Demonstrate the capability to perform the objective within 4 hours in 100% of the identified requirements.
	b	Pull data on a schedule from disparate data sources	Demonstrate the capability to perform the objective within 1 hour in 100% of the identified requirements.
	c	Conduct data cleansing and initial profiling as appropriate	Demonstrate the capability to perform the objective within 2 hours in 100% of the identified requirements.
	d	Write the data into a data warehouse	Demonstrate the capability to perform the objective within 1 hour in 100% of the identified requirements.
	e	Create a metadata repository with full bi-directional linkages with all Data Integration, Reporting, Data Visualization and Advanced Analytics components	Demonstrate the capability to perform the objective within 1 hour in 100% of the identified requirements.
2		Conduct analysis on the data available. This includes indexing and profiling, integrated data and text mining, predictive modeling, forecasting and optimization as required to meet mission requirements	Demonstrate the capability to perform the objective within 2 hours in 100% of the identified requirements.
3		Develop networks , along with soft and hard links, based on the data provided	Demonstrate the capability to perform the objective within 4 hours in 100% of the identified requirements.
4		Score the networks as benign or suspicious based on criteria established by the PM	Demonstrate the capability to perform the objective within 2 hours in 100% of the identified requirements.
5		Identify and list key network behaviors and potential vulnerabilities	Demonstrate the capability to perform the objective within 2 hours in 100% of the identified requirements.

6		Push information out to end users in multiple formats (portal, PDA, reports, alerts, etc) based on responsibilities, roles and access rights	Within 4 hours of completion of analysis and upon release by the appropriate authority
7		Provide standard interactive, parameterized “What if” interfaces for end users based upon model outcomes and parameters.	Interfaces updated with latest models and parameters within 4 hours of completion of analysis and upon release by the appropriate authority
8		Provide controlled data access via system metadata layer for ad-hoc, reporting, data visualization, and advanced analytic modeling to end users based on responsibilities roles and access rights.	Within 4 hours of completion of analysis and upon release by the appropriate authority
9		Provide integrated accuracy monitoring of “predicted” versus “verified” condition monitoring of scoring outcomes with threshold triggers to recalibrate or redevelop scoring algorithms.	Demonstrate the capability to perform the objective within 2 hours of the completion of scoring activities.

4.2. Key Performance Parameters (KPPs)

All Operational Performance Parameters are considered mandatory.

4.3 System Performance.

4.3.1 Mission Scenarios

The purpose of a solution is to provide analysts, investigators and decision makers the ability to gain more thorough and detailed insight of cross-agency data, using proven COTS technology, to better identify potential threats to the United States.

The solution shall be deployable into a cross-agency headquarters level environment operating at a minimum SECRET classified level. The three primary purposes of a solution are to:

- Extract, cleanse and integrate data from Federal, state, local and tribal agencies into a data repository, staged for the application of advanced analytics. Data must be accessed from a range of transactional, operational and individual sources in a variety of software platforms residing in different environments with operating systems with various levels of classification, potentially worldwide.

- Apply advanced analytics to develop data-driven intelligence on social, operational and financial networks potentially involved in domestic IED/VB-EID/suicide bomb attacks on the United States. The term "advanced analytics" is interpreted to include but is not limited to the use of text and data mining, predictive modeling, forecasting and optimization, within the context of innovative thought leadership to develop the most comprehensive and integrated understanding of a given threat, as well as how to validate and respond to it.

- Communicate the analytic results to decision makers, analysts and investigators within the appropriate cross-departmental environment for action. Information and intelligence should be available in a variety of outputs. Access and permissions must be based on roles and responsibilities.

4.3.2 System Performance Parameters

The Performance Parameters are addressed above in Section 4.1

4.3.3 Interoperability

- Interoperability, defined as the capability of applications to exchange information and to operate cooperatively using this information, is a critical aspect of this solution, since data will be integrated from disparate sources through the Federal, state, local and tribal infrastructure. The solution must sit atop the systems identified by the department/agency and extract, cleanse and load to a data warehouse or repository (or potentially multiple repositories) with minimal human interaction and full transparency, the ability to audit, read and write in native language to the source systems is also required. A solution shall reduce overhead and make access to the required sources and data elements efficient and timely.

4.3.4 Human Interface Requirements

A solution shall operate in a controlled, IT-type environment at a department/agency owned or contracted facility within the United States.

The solution shall comply with all Federal core configuration requirements.

4.3.5 Logistics and Readiness

- A solution shall include both production and backup architecture to ensure the solution maintains at least a 99% rate with a ≤ 3 hours resumption of service capability in case of catastrophic failure.

- Logistics (maintenance and supply) requirements are addressed below in Sections 5.1 and 5.2

5. System Support

5.1 Maintenance

Maintenance of the hardware for the hosting system shall be the responsibility of the department/agency. Maintenance of the technology platform, including resolution of technical support issues and installation of upgrades or fixes, will be the responsibility of the solution provider.

5.2 Supply

The operating environment—including a combination of separate development, test, production and backup environments—will be operated and maintained by the department/agency with solution-specific supplies or spares for sustained operation with 99% availability.

A solution will include both system and agency-specific solution documentation, tailored for agency use and delivered both online and as standalone media (CD/DVD or thumb drive), in required.

5.3 Support Equipment

A solution shall require support equipment that is readily available as commercial-off-the-shelf equipment.

5.4 Training

A solution shall provide for the full range of training needed to operate the system and provide the services required at Technology Readiness Level (TRL) 9. A solution will specify the specific training provided to ensure that users are capable of operating and using a proposed system. It will identify the training required for each component of the overall solution along with metrics to verify that each person participating in the training is certified upon completion.

A solution will make use of online and self-paced learning modes, but, where appropriate, provide for classroom training. Classroom training should be tailored to the needs of the agency-specific solution and be conducted training facilities provided by the agency.

Training materials will be provided, in electronic format—online or via portable media—for all courses, including online or distance learning modules.

5.5 Transportation and Facilities

A solution will be hosted in a government-specified, controlled environment with no anticipated requirement for transportation.

6. Force Structure

A solution must consist of COTS software deployed in development, test, production and backup environments, with the required hardware sets provided by the government. The development and test environments will be used to tailor a production-ready, COTS technology platform to ensure that the environment available to the end user community is at TRL 9. The environments could be separately located.

The system shall be certified for use at least at the SECRET classified level and operate in both the unclassified and classified environments with the appropriate safeguards in place.

The solution should scale to accept classified and unclassified data feeds from an unlimited number of sources, regardless of the host operating system or base application. In addition, the solution should accommodate manual input from unlimited number of users via the Internet or WAN/LAN, as well as an unlimited number of end user via the same connectivity.

7. Schedule

A solution will be at TRL 9 within 6 months from the day access to the identified data sources is confirmed by the agency Program Manager and the solution provider. TRL 9 is generally defined as the ability to do the following:

- Link disparate, cross-agency data sources and integrate required elements into a data warehouse/repository with a supporting metadata layer

- Enhance data consistency and accuracy using data quality/cleansing software

- Use data/text mining, predictive modeling and other advanced analytics methodology to provide insight to relevant data

- Identify suspicious social, financial and operational networks that may be appropriate for further analysis or investigation

• Push actionable information out to decision makers and investigators via the Internet, LAN/WAN, email, etc.

8. System Affordability

The hardware will be provided by the department/agency and the software component of the solution shall include all technology required for:

- Data integration, cleansing storage and management
- Analytics, including data and text mining, predictive modeling and forecasting
- Development and scoring of social networks
- Deployment of information to end users with a capability to run stored processes from within Microsoft Office desktop applications, do ad hoc query and analysis and drill down
- Full control of access to the solution by a local system administrator
- Full documentation at the system administrators, analysts and end users levels
- 24/7 Technical support

Training shall be included for system administrators, analysts and end users in Web-deployed and on-site classroom training packages.

The services component of the solution shall include:

- Installation and configuration of all software, including all fixes and upgrades
- Development of analytical models
- Scoring and linking within the context of the social network models
- Knowledge Transfer to government employees or contract personnel as directed

The total delivered price should be ≤ $500,000 (includes department/agency usage rights). In addition, provide a fixed price proposal for seat or facility licensing fee for users outside of the department/agency.

A conservative estimate of the potential available market is over 250,000 seats in the United States alone. It is conservatively estimated that there are more than 5 Federal departments/agencies identified as potential users for the proposed system.