

Report to Congress regarding the Terrorism Information Awareness Program

In response to Consolidated Appropriations Resolution, 2003, Pub. L.
No. 108-7, Division M, § 111(b)

May 20, 2003

Report to Congress regarding the Terrorism Information Awareness Program

In response to Consolidated Appropriations Resolution, 2003, Pub. L.
No. 108-7, Division M, § 111(b)

Executive Summary

May 20, 2003

Terrorism Information Awareness Program

Preface

The Consolidated Appropriations Resolution, 2003, Pub. L. No. 108-7, Division M, § 111(b) provides for the submission of a report to Congress, within 90 days of the President's signing the law, regarding the Total Information Awareness program, now called the Terrorism Information Awareness (TIA) program, a Defense Advanced Research Projects Agency (DARPA) research and development program initiated in the aftermath of the September 11, 2001 terrorist attacks on New York and Washington.

Executive Summary

The Defense Advanced Research Projects Agency (DARPA) is charged with conducting research and development for the Department of Defense (DoD). By doing so, DARPA furnishes DoD with leading-edge technologies to help the department execute its critical national security mission. DARPA often produces prototype systems for conducting experiments that address the urgent needs of DoD. If successful and as appropriate, such prototype systems would be transitioned into operational use by executing agencies of the government.

Terrorism Information Awareness (TIA)¹ is such a prototype system/network. It is a research and development program that will integrate advanced collaborative and decision support tools; language translation; and data search, pattern recognition, and privacy protection technologies into an experimental prototype network focused on combating terrorism through better analysis and decision making. If successful, and if deployed, this program of programs would provide decision- and policy-makers with advance actionable information and knowledge about terrorist planning and preparation activities that would aid in making informed decisions to prevent future international terrorist attacks against the United States at home or abroad. In short, DoD's aim in TIA is to seek to make a significant leap in technology to help those working to "connect the dots" of terrorist-related activity. A TIA-like system/network could provide the defense and intelligence communities with tools and methods to solve many of the problems that have been identified in the aftermath of the attacks against the United States on September 11, 2001,² and that are related to improving information analysis in our continuing war against terrorism.

¹ Previously known as Total Information Awareness, this name created in some minds the impression that TIA was a system to be used for developing dossiers on U.S. citizens. That is not DoD's intent in pursuing this program. Rather, DoD's purpose in pursuing these efforts is to protect U.S. citizens by detecting and defeating foreign terrorist threats before an attack. To make this objective absolutely clear, DARPA has changed the program name to Terrorism Information Awareness.

² Final Report of the Joint SSCI/HPSCI Inquiry into the Events of 9/11/01 dated Dec 10, 2002.

DoD's TIA research and development is aimed at providing capabilities to users/analysts/operators to address a perennial array of problems that have beset analysis of complex threats, including sharing data across agency boundaries and exploiting both classified and unclassified information, in a more systematic fashion.

These problems exist in part because of a lack of applied technology to aid the human processes. Today, the amount of information that needs to be considered far exceeds the capacity of the unaided humans in the system. Adding more people is not necessarily the solution. DoD believes that there is a need to provide a much more systematic, methodological approach that automates many of the lower-level data manipulation tasks that can be done well by machines guided by human users. Such an approach would, in turn, allow users more time for higher-level analysis that depends critically on a human's unique cognitive skills.

TIA is one of several research and development programs in DARPA's Information Awareness Office (IAO), which was established in January 2002. In the aftermath of the September 11 terrorist attacks, DARPA formed IAO in part to bring together, under the leadership of one technical office director, several existing DARPA programs focused on applying information technology to combat terrorist threats. DARPA also recognized that new programs would be needed to fully address the technology requirements of a complete prototype system/network to respond to the particular demands of the terrorist threat. DARPA envisions TIA as the system/network-level integration program while other IAO programs are designed to furnish technologies and components that compose the overall program. As conceived by DARPA, TIA would integrate these technologies and provide some or all of them to various organizations for experiments, while assessing the system's utility in various operationally relevant contexts.

The TIA research and development program began in FY 2003. Funding for FY 2003 through FY 2005 as proposed in the FY 2004 President's Budget submission is \$53,752,000. A number of organizations in the DoD and Intelligence Community have shown great interest in working with the TIA program to test and evaluate technologies.

DARPA provides a system/network infrastructure and concepts; software analytical tools; software installation; training; software performance evaluation; and integration and evaluation of user comments on modifications and additions to the software. Participating organizations from DoD and the Intelligence Community provide facilities and personnel to evaluate these products and use data currently available to them under existing laws, regulations and policies.

Five major investigation threads are currently being pursued as a part of TIA and are driving much of the development and experimental activity in the TIA program. These five threads are: secure collaborative problem solving, structured discovery with security, link and group understanding, context aware visualization, and decision making with corporate memory.

- **Secure Collaborative Problem Solving.** A collaborative environment is sought that would enable ad hoc groups to quickly form within and across agency boundaries to bring relevant data, diverse points of view, and experience together to solve the complex problems associated with countering terrorism.

- **Structured Discovery with Sources and Methods Security.** A wide range of intelligence data, both classified and open source, may need to be searched to find relevant information for understanding the terrorist intent. DARPA believes that to have any hope of making sense of this wide range of data, a more structured and automated way of approaching the problem is needed.
- **Link and Group Understanding.** One of the characteristics of the terrorist threat is that terrorist organizational structures are not well understood and are purposefully designed to conceal their connections and relationships. IAO is researching software that can discover linkages among people, places, things, and events related to possible terrorist activity.
- **Context Aware Visualization.** DARPA believes that better ways are needed to visualize information than text-based lists, tables, and long passages of unstructured text. Such visualization concepts should respond to a broad range of potential users with wholly different roles and responsibilities.
- **Decision Making with Corporate Memory.** Decision-makers must consider a full range of possible options to deal with complex asymmetric threats, particularly in light of rapidly changing circumstances. DARPA's activities in this area are premised on the view that understanding how certain decisions played out in the past is critical to formulating current decision options.

The TIA program is a research and development project. The program is integrating and testing information technology tools. DARPA affirms that TIA's research and testing activities are only using data and information that is either (a) foreign intelligence and counter intelligence information legally obtained and usable by the Federal Government under existing law, or (b) wholly synthetic (artificial) data that has been generated, for research purposes only, to resemble and model real-world patterns of behavior .

The Department of Defense, which is responsible for DARPA, has expressed its full commitment to planning, executing, and overseeing the TIA program in a manner that protects privacy and civil liberties. Safeguarding the privacy and the civil liberties of Americans is a bedrock principle. DoD intends to make it a central element in the Department of Defense's management and oversight of the TIA program.

The Department of Defense's TIA research and development efforts address both privacy and civil liberties in the following ways:

- The Department of Defense must fully comply with the laws and regulations governing intelligence activities and all other laws that protect the privacy and constitutional rights of U.S. persons.
- As an integral part of its research, the TIA program itself is seeking to develop new technologies that will safeguard the privacy of U.S. persons.

- TIA's research and testing activities are conducted using either real intelligence information that the federal government has already legally obtained, or artificial synthetic information that, ipso facto, does not implicate the privacy interests of U.S. persons.

The report does not recommend any changes in statutory laws, but instead contemplates that any deployment of TIA's search tools may occur only to the extent that such a deployment is consistent with current law. Accordingly, the report specifically notes that the strictures of current law protecting certain categories and sources of information may well constrain or (as a logistical matter) completely preclude deployment of TIA search tools with respect to such data.

Moreover, to the extent that TIA research and development technology is ever applied to data sources that contain information on U.S. persons, the privacy issues raised by these tools are significant ones that will require careful and serious examination. Because TIA is still largely in the research stage, any analysis of these issues is necessarily tentative and preliminary. Several factors would need to be considered in evaluating TIA's suitability for deployment in particular contexts.

- The *efficacy and accuracy* of TIA's search tools must be stress-tested and demonstrated. The tools must be shown to be sufficiently precise and accurate – i.e., a search query results in *only* that information that is responsive to the query. DARPA has expressed its commitment to the necessary testing to ensure the technological accuracy of TIA's search tools.
- It is critical that there be *built-in operational safeguards* to reduce the opportunities for abuse. DARPA is already researching whether and how it may be able to build in controls that, at an architectural level, would govern the TIA program tools. Among the controls being researched are automated audit trails to document who accessed the system and how it was used during the session; anonymization of sources of data and of the persons mentioned in the underlying data, so that these data could not be revealed unless it is lawful and warranted; selective revelation of data, so that additional permissions would need to be obtained in order to receive additional data; and rigorous access controls and permissioning techniques. TIA's ultimate suitability for particular purposes will depend heavily upon DARPA's success on these technological issues.
- It will also be essential to ensure that *substantial security measures* are in place to protect these tools from unauthorized access by hackers or other intruders. Some of these measures must be built-in at the architectural level; others will involve the adoption of policies that prescribe who may have access, for what purposes, and in what manner.
- Any agency contemplating deploying TIA tools for use in particular contexts will be required first to conduct a *pre-deployment legal review*. In this regard, the DoD General Counsel has directed each operational component within DoD that hosts TIA technologies to prepare a substantive legal review that examines the relationship

between that component and TIA, and analyzes the legal issues raised by the underlying program to which the TIA tools will be applied. The General Counsel has advised that all such relationships should be documented in a memorandum of agreement to ensure the relationship is clearly understood by all parties. The DCI's General Counsel is taking comparable steps with respect to elements of the Intelligence Community, and the Department of Justice would do so if it ever decides to deploy any TIA technology.

- There will be a need for any user agency to adopt policies establishing *effective oversight* of the actual use and operation of the system before it is deployed in particular contexts. There must be clear and effective accountability for misuse of the system.

As DARPA endeavors to achieve these technological developments, the Secretary of Defense will, as an integral part of oversight of TIA research and development, continue to assess emerging potential privacy and civil liberties impacts through an oversight board composed of senior representatives from DoD and the Intelligence Community, and chaired by the Under Secretary of Defense (Acquisition, Technology and Logistics). The Secretary of Defense will also receive advice on legal and policy issues, including privacy, posed by TIA research and development from a Federal Advisory Committee composed of outside experts.

The Department of Defense has expressed its intention to address privacy and civil liberties issues squarely as they arise, in specific factual and operational contexts and in full partnership with other Executive Branch agencies and the Congress. The protection of privacy and civil liberties is an integral and paramount goal in the development of counterterrorism technologies and in their implementation. If these technologies can be developed, the privacy and civil liberties issues noted above would have to be carefully considered and resolved in advance of deployment.

Report to Congress regarding the Terrorism Information Awareness Program

In response to Consolidated Appropriations Resolution, 2003, Pub. L.
No. 108-7, Division M, § 111(b)

Detailed Information

May 20, 2003

TABLE OF CONTENTS

<i>Program Information</i>	1
DARPA’s Information Awareness Office	1
TIA and High-Interest TIA-Related Program Information	3
Terrorism Information Awareness (TIA)	3
Genisys	5
Genisys Privacy Protection	6
Evidence Extraction and Link Discovery (EELD)	7
Scalable Social Network Analysis (SSNA)	9
MisInformation Detection (MInDet)	9
Human Identification at a Distance (HumanID) Program	10
Activity, Recognition and Monitoring (ARM)	11
Next-Generation Face Technology (NGFR)	12
<i>TIA Efficacy</i>	13
The Promise of TIA	13
How TIA Would Work	14
Measuring TIA Progress and Effectiveness	15
Status of Component Research	17
<i>Laws and Regulations Governing Federal Government Information Collection</i>	18
<i>TIA’s Impact on Privacy and Civil Liberties, and Recommended Practices, Procedures, Regulations or Legislation for TIA Deployment and Implementation to Eliminate or Minimize Adverse Effects</i>	27
Overview	27
Relevant Information Privacy Principles	28
Preliminary Assessment of Privacy Implications of TIA and Pertinent Recommendations	30
<i>Appendix A – Detailed Description of TIA and High-Interest TIA-Related Programs</i>	A-1
Terrorism Information Awareness (TIA)	A-1
Genisys	A-10
Genisys Privacy Protection	A-12
Evidence Extraction and Link Discovery (EELD)	A-14
Scalable Social Network Analysis (SSNA)	A-16
MisInformation Detection (MInDet)	A-17
Human Identification at a Distance (HumanID) Program	A-18
Activity, Recognition, and Monitoring (ARM)	A-21
Next-Generation Face Technology (NGFR)	A-22
<i>Appendix B – Other IAO Programs</i>	B-1
Genoa II	B-1
Wargaming the Asymmetric Environment (WAE)	B-5
Rapid Analytical Wargaming (RAW)	B-7
Futures Markets Applied to Prediction (FutureMAP)	B-8
Automated Speech and Text Exploitation in Multiple Languages	B-9
Effective, Affordable, Reusable Speech-to-Text (EARS)	B-10
Translingual Information Detection, Extraction, and Summarization (TIDES)	B-12
Global Autonomous Language Exploitation (GALE)	B-16

Situation Presentation and Interaction	B-17
Babylon	B-17
Symphony	B-20
Bio-Event Advanced Leading Indicator Recognition Technology	B-21
(Bio-ALIRT)	B-21

Appendix C – Information Paper on Intelligence Oversight of INSCOM’s Information Operations Center (IOC)	C-1
---	------------

Appendix D – TIA Program Directives	D-1
--	------------

Appendix E – DARPA–U.S. Army INSCOM Memorandum of Agreement	E-1
--	------------

FIGURES

Figure 1 - IAO Organization	2
Figure 2 - TIA Reference Model	A-7

Program Information

DARPA's Information Awareness Office

Since 1996, the Defense Advanced Research Projects Agency (DARPA) has been developing information technologies to counter asymmetric threats. Although the individual efforts attacked significant pieces of the problem, they lacked an integrated approach. September 11, 2001, brought home the need for a new research focus on counterterrorism. Already in possession of individual pieces of the counterterrorism puzzle, DARPA created the Information Awareness Office (IAO) in January 2002 to integrate advanced technologies and accelerate their transition to operational users. The relevant existing programs were moved to this new technical office, and some new programs were started in FY 2003. About the same time, the U.S. Army Intelligence and Security Command (INSCOM) was developing the Information Dominance Center (now titled the Information Operations Center). Discussions between DARPA and INSCOM resulted in a joining of forces to create a unique environment for research and development (R&D) to directly and immediately enhance the capabilities of intelligence analysts grappling with ongoing real-world threats. DoD believes this will help ensure transition of the R&D programs to eventual operational use and respond to the urgency of problem solutions. The events of September 11 heightened awareness of the increasing frequency, complexity, and lethality of these threats. In response, the IAO is directing a portfolio of R&D programs focused on significantly improving counterterrorism capabilities in DoD and other agencies within the greater Intelligence Community.

The organization of IAO is shown in the following Figure 1. IAO is one of eight technical offices under the leadership and management of the Director of DARPA. The mission statement for IAO states in part:

The DARPA Information Awareness Office (IAO) will imagine, develop, apply, integrate, demonstrate and transition information technologies, components, and prototype closed-loop information systems that will counter asymmetric threats by achieving total information awareness useful for preemption, national security warning, and national security decision making.

There are two major sections of the IAO. One section (left side of diagram) shows the technology side of the office which is organized by programs that develop technologies and components. Each program is led by a program manager who has contracts with universities, commercial companies, and government laboratories to perform the actual R&D. Technologies and components from all these programs (except Babylon and Symphony) may be provided to the Terrorism Information Awareness (TIA) effort, which is the system-level effort (right side of diagram). These programs are supplemented with components from other government programs and commercial sources where appropriate and necessary to create early versions of a prototype system. In the TIA R&D program, a prototype network has been established for integrating and testing tools and concepts in an operational environment. The main node of TIA network is located in the INSCOM Information Operations Center. Additional TIA network nodes are located at subordinate INSCOM commands and other participating organizations from DoD and the Intelligence Community. DARPA affirms that these agencies and commands are using data that is available to them under existing laws and procedures for the tests.

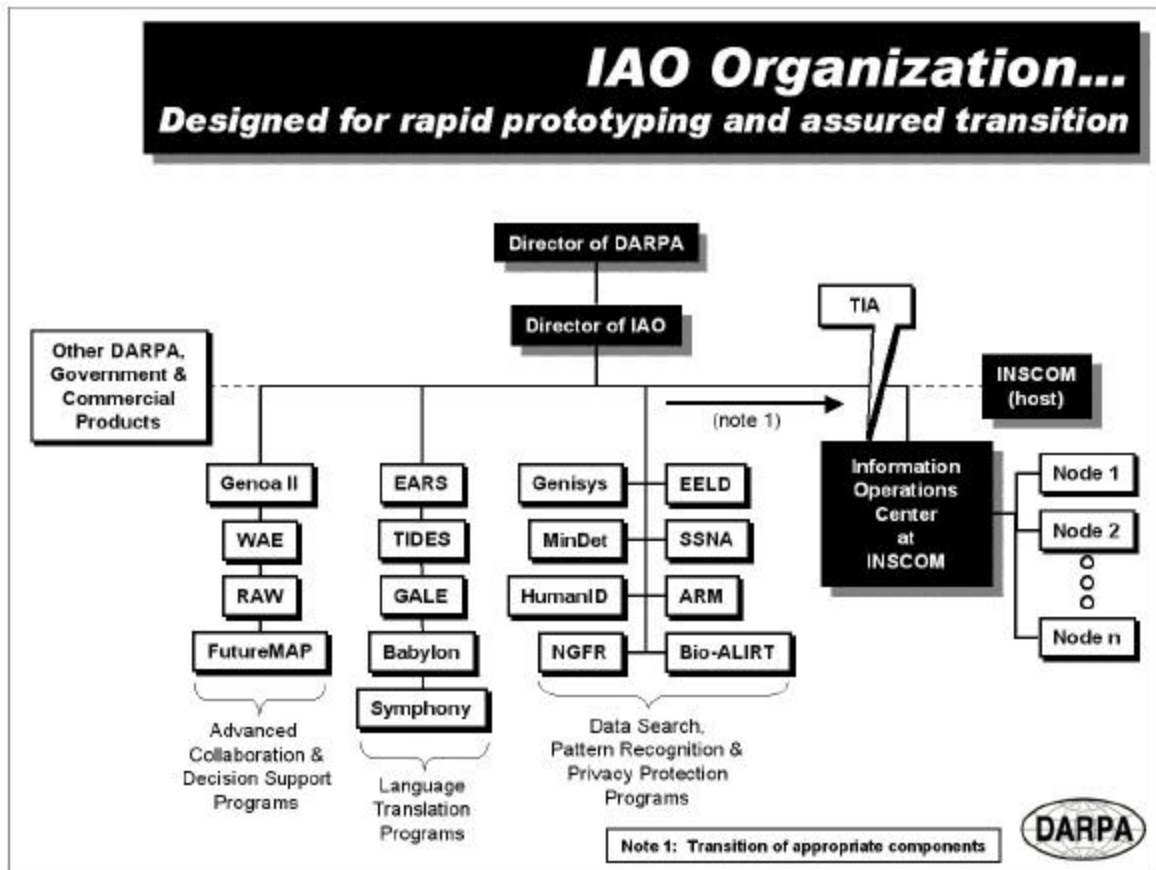


Figure 1 - IAO Organization

The R&D being conducted in these programs can be divided into four categories:

- **Technology Integration and Experimentation Programs**
 - Terrorism Information Awareness (TIA)
- **Advanced Collaborative and Decision Support Programs**
 - Genoa II (collaboration and decision support)
 - Wargaming the Asymmetric Environment (WAE)
 - Rapid Analytical Wargaming (RAW)
 - Futures Markets Applied to Prediction (FutureMAP)
- **Language Translation Programs**
 - Effective, Affordable, Reusable Speech-to-Text (EARS)
 - Translingual Information Detection, Extraction and Summarization (TIDES)
 - Global Autonomous Language Exploitation (GALE)
 - Babylon (natural language two-way translation for military field operations)
 - Symphony (natural language human-to-computer interface for field operations)

- **Data Search, Pattern Recognition, and Privacy Protection Programs**
 - Genisys (data base access, data repository, and privacy protection)
 - Evidence Extraction and Link Discovery (EELD)
 - Scalable Social Network Analysis (SSNA)
 - MisInformation Detection (MInDet)
 - Bio-Event Advanced Leading Indicator Recognition Technology (Bio-ALIRT)
 - Human Identification at a Distance (HumanID) Program
 - Activity, Recognition, and Monitoring (ARM)
 - Next-Generation Face Recognition (NGFR)

This report addresses for TIA and high-interest TIA-related programs:

- Program overview
- Program schedule
- FY 2004 President's Budget

TIA research and development and high-interest TIA-related programs are discussed in further detail in Appendix A, which provides each program's technical approach, relationship to TIA, and program transition/deployment plans. The high-interest TIA-related programs, those programs involving data access, data search, pattern recognition and privacy protection, are those that are deemed relevant to any discussion of technologies which, if applied to data on US persons, would raise serious issues about privacy. These programs are: TIA, Genisys, Genisys Privacy Protection, EELD, SSNA, MInDet, HumanID, ARM, and NGFR. The details of other IAO programs are included in Appendix B for completeness.

TIA and High-Interest TIA-Related Program Information

Note: The target date for the deployment of each project listed in this report is the completion date listed, unless identified differently in the descriptive paragraphs.

Terrorism Information Awareness (TIA)

The TIA research and development program aims to integrate information technologies into a prototype to provide tools to better detect, classify, and identify potential foreign terrorists. TIA's research and development goal is to increase the probability that authorized agencies of the United States can preempt adverse actions.

The TIA research and development efforts seek to integrate technologies developed by DARPA (and elsewhere, as appropriate) into a series of increasingly powerful prototype configurations that can be stress-tested in operationally relevant environments using real-time feedback to refine concepts of operation and performance requirements down to the technology component level. In a sense, TIA is a program of programs whose goal is the creation of a counterterrorism information architecture that would:

- Increase the information coverage by an order-of-magnitude via access and sharing that can be easily scaled.
- Provide focused warnings within an hour after a triggering event occurs or an articulated threshold is passed.
- Automatically cue analysts based on partial pattern matches and has patterns that cover at least 90 percent of all known previous foreign terrorist attacks.
- Support collaboration, analytical reasoning, and information sharing so analysts can hypothesize, test, and propose theories and mitigating strategies about possible futures, thereby enabling decision-makers to effectively evaluate the impact of current or future policies.

DARPA will work in close collaboration with other participating organizations from DoD and the Intelligence Community for TIA research and development evaluation, technology maturation, and possible transition partners. In the near-term, the main effort will take place within the U.S. Army Intelligence and Security Command (INSCOM). Using output from other programs in IAO, other government programs, and commercial products, the TIA Program intends to create fully functional, integrated, leave-behind component prototypes that are reliable, easy to install, and packaged with documentation and source code (though not necessarily complete in terms of desired features) that will enable the Intelligence Community to evaluate new TIA technology through experimentation and rapidly transition it to operational use, as appropriate.

DoD, on its own, has taken several measures in an effort to ensure that TIA research and development program managers and performing contractors are acutely aware of the unique R&D environment at INSCOM and the special requirements for properly handling sensitive data in such a setting. See Appendix C, “Information Paper on Intelligence Oversight of INSCOM’s Information Operations Center (IOC)”; Appendix D, “TIA Program Directives”; and Appendix E, “DARPA-U.S. Army INSCOM Memorandum of Agreement.”¹ DoD reaffirms its commitment to ensuring that TIA Program activities are conducted in full compliance with relevant policies, laws, and regulations, including those governing information about U.S. persons.

TIA PROGRAM - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$9,233	\$20,000	\$24,519	FY 2007

¹ DARPA intends to use the memorandum of agreement in Appendix E as a model to support the establishment of additional TIA test nodes.

PROGRAM SCHEDULE: TIA began in FY 2003. The current schedule through FY 2005 follows.

Milestone	FY/Quarter
Design and develop an initial TIA system architecture and document in a system design document.	FY03 (1Q)
Develop, integrate, and deploy initial TIA system prototype based on a suite of COTS, GOTS, and various analytical and collaborative software tools from several IAO programs (i.e., Genoa, TIDES, EELD).	FY03 (1Q)
Establish a baseline-distributed infrastructure consisting of software, hardware, and users to support end-to-end continuous experiment environment for TIA system technology.	FY03 (1Q)
Submit TIA system performance measurement processes and metrics (v1.0).	FY03 (2Q)
Initial review Phase II metrics.	FY03 (4Q)
Final exam and transition to info-cockpit prototype design.	FY03 (4Q)
Midterm exam – metrics.	FY03 (4Q)
Plan and execute threat-based red teaming experiments spanning various types of terrorist attacks, CONOPS, and information signals.	FY03-FY04
Apply TIA system technology using real-world data and real users to solve real-world problems.	FY03-FY05
Identify and assess emerging information technology and CONOPS for use in TIA system network infrastructure and for analytical tools.	FY03-FY05
Explore concepts and techniques for analyzing and correlating new data sources applicable to counter terrorism.	FY03-FY05
Develop enhanced TIA system prototypes, metrics, and experiments.	FY03-FY05
Harden and mature fragile TIA system technology and corresponding CONOPS successfully demonstrated within experiments.	FY04-FY05

Genisys

The Genisys Program seeks to produce technology for integrating and broadening databases and other information sources to support effective intelligence analysis aimed at preventing terrorist attacks on the citizens, institutions, and property of the United States. DARPA's goal is to make databases easy to use so users can increase the level of information coverage, get answers when needed, and share information among agencies faster and easier. DARPA believes that, in order to predict, track, and thwart attacks, the United States needs databases, containing information

about potential terrorists and possible supporters, terrorist material, training/preparation/rehearsal activities, potential targets, specific plans, and the status of our defenses. In DARPA’s view, current commercial technology is far too complex and inflexible to easily integrate relevant existing databases or to create new databases for systems that collect legally obtained data in paper and unstructured formats. DARPA’s premise is that information systems need to be easier to use; thus, technologies must be more sophisticated.

DARPA’s vision is that Genisys technologies will make it possible for TIA properly to access the massive amounts of data on potential foreign terrorists. In FY 2003, the program aims to develop a federated database architecture and algorithms that would allow analysts and investigators to more easily obtain answers to complex questions by eliminating their need to know where information resides or how it is structured in multiple databases. In FY 2004, the program aims to create technology for effectively representing and resolving uncertainty and inconsistency in the data values so that intelligence analysis will be faster and more certain.

GENISYS - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$6,964	\$7,241	\$8,588	FY 2007

PROGRAM SCHEDULE: The Genisys Program began in FY 2003 and will conclude in FY 2007.

Milestone	FY/Quarter
Develop abstract schema (Phase I).	FY03 (4Q)
Create technology for effectively representing uncertainty in the database (Phase II).	FY04 (4Q)
Develop virtually centralized databases with no practical size limit (Phase III).	FY05 (4Q)
Improve performance and transition (Phase IV).	FY07 (4Q)

Genisys Privacy Protection

The Genisys Privacy Protection Program aims to create new technologies to ensure personal privacy in the context of improving data analysis for detecting, identifying, and tracking terrorist threats. Information systems and databases have the potential to identify terrorist signatures through the transactions they make, but Americans are rightfully concerned that data collection, integration, analysis, and mining activities implicate privacy interests. The Genisys Privacy Protection Program aims to provide *security with privacy* by providing certain critical data to analysts while controlling access to unauthorized information, enforcing laws and policies through software mechanisms, and ensuring that any misuse of data can be quickly detected and addressed. Research being conducted under other IAO programs may indicate that information about terrorist planning and preparation activities exists in databases that also contain

information about U.S. persons. Privacy protection technologies like those being developed under the Genisys Privacy Protection Program would be essential to protect the privacy of U.S. citizens should access to this sort of information ever be contemplated. In FY 2003, DARPA aims to develop algorithms that prevent unauthorized access to sensitive identity data based on statistical and logical inference control, and create roles-based rules to distinguish between authorized and unauthorized uses of data and to automate access control. In FY 2004, DARPA will seek to enhance these algorithms and provide an immutable audit capability so investigators and analysts cannot misuse private data without being identified as the culprits. These technologies are also applicable to protecting intelligence methods and sources and reducing the potential “insider threat” in intelligence organizations.

GENISYS PRIVACY PROTECTION - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$3,921	\$3,982	\$5,900	FY 2007

PROGRAM SCHEDULE: The Genisys Privacy Protection Program began in FY 2003 and will conclude in FY 2007. The current schedule through FY 2006 follows.

Milestone	FY/Quarter
Create privacy algorithms (Phase I).	FY03 (4Q)
Create a trusted guard for safeguarding the personal privacy of U.S. citizens (Phase II).	FY04 (4Q)
Develop algorithms for automating audit and detecting privacy violations (Phase III).	FY06 (4Q)

Evidence Extraction and Link Discovery (EELD)

The objective of the EELD program is to develop a suite of technologies that will automatically extract evidence about relationships among people, organizations, places, and things from unstructured textual data, such as intelligence messages or news reports, which are the starting points for further analysis. In DARPA’s view, this information can point to the discovery of additional relevant relationships and patterns of activity that correspond to potential terrorist events, threats, or planned attacks. These technologies would be employed to provide more accurate, advance warnings of potential terrorist activities by known or, more important, unknown individuals or groups. DARPA believes that they will allow for the identification of connected items of information from multiple sources and databases whose significance is not apparent until the connections are made. To avoid needless, distracting, and unintended analysis of ordinary, legitimate activities, these technologies seek to ensure that intelligence analysts view information about only those connected people, organizations, places, and things that are of interest and concern and that require more detailed analysis.

In FY 2002, the EELD Program demonstrated the ability to extract relationships in several sets of text; the ability to distinguish characteristic, relevant patterns of activity from similar legitimate activities; and improvements in the ability to classify entities correctly based on their connections to other entities. These advances have been applied to significant intelligence problems. In FY 2003, the diversity of detectable relationships is being increased, the complexity of distinguishable patterns is being increased, and the ability to automatically learn patterns will be demonstrated. In FY 2004, the program will evaluate and transition selected components to the emerging TIA network nodes in the Defense and intelligence communities and will integrate the ability to learn patterns of interest with the ability to detect instances of those patterns. In summary, the EELD Program seeks to develop technology not only for “connecting the dots,” but also for deciding which dots to connect—starting with suspect people, places, or organizations known or suspected to be suspicious based on intelligence reports; recognizing patterns of connections and activity corresponding to scenarios of concern between these people, places, and organizations; and learning patterns to discriminate as accurately as possible between real concerns and apparently similar but actually legitimate activities.

EELD - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$12,309	\$16,552	\$10,265	\$5,515	FY 2005

PROGRAM SCHEDULE: The EELD effort began in FY 2001 and will conclude in FY 2005.

Milestone	FY/Quarter
Develop Test Set	FY02 (1Q)
1st Extraction Evaluation	FY02 (4Q)
1st Link Discovery Evaluation	FY02 (4Q)
1st Pattern Learning Evaluation	FY02 (4Q)
2nd Extraction Evaluation	FY03 (3Q)
2nd Link Discovery Evaluation	FY03 (4Q)
2nd Pattern Learning Evaluation	FY03 (4Q)
Integrated Extraction Module	FY03 (4Q)
Integrated Link Discovery Module	FY04 (3Q)
Integrated Pattern Learning Module	FY05 (2Q)
Classified Evaluation	FY04 (3Q)
Final Evaluation	FY05 (3Q)

Scalable Social Network Analysis (SSNA)

The purpose of the SSNA algorithms program is to extend techniques of social network analysis to assist with distinguishing potential terrorist cells from legitimate groups of people, based on their patterns of interactions, and to identify when a terrorist group plans to execute an attack. Current techniques in social network analysis take into account only a link among individuals without characterizing the nature of the connection. DARPA believes that there is a need to simultaneously model multiple connection types (e.g., social interactions, financial transactions, and telephone calls) and combine the results from these models. DARPA also believes that there is a need to analyze not only a single “level,” such as connections between people or between organizations, but multiple “levels” simultaneously, such as interactions among people and the organizations of which they are a part. Based on publicly available information about the September 11 hijackers, contractors working under the EELD Program and Small Business Innovation Research (SBIR) contracts have demonstrated the feasibility of using these techniques to identify the transition of terrorist cell activity from dormant to active state by observing which social network metrics changed significantly and simultaneously.

In FY 2003, DARPA will develop a library of models of social network features that represent potential terrorist groups. In FY 2004, DARPA will develop algorithms that allow for the discovery of instances of these models in large databases.

SSNA - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u> \$000	<u>FY 2003</u> \$000	<u>FY 2004</u> \$3,348	<u>FY 2005</u> \$4,040	<u>Completion Date</u> FY 2007
-------------------------	-------------------------	---------------------------	---------------------------	-----------------------------------

PROGRAM SCHEDULE: SSNA begins in FY 2004 and concludes in FY 2007. A milestone schedule is under consideration.

MisInformation Detection (MInDet)

The purpose of the MInDet Program is to reduce DoD vulnerability to open source information operations by developing the ability to detect intentional misinformation and to detect inconsistencies in open source data with regard to known facts and adversaries’ goals. As a new program, MInDet seeks to improve national security by permitting the intelligence agencies to evaluate the reliability of a larger set of potential sources and, therefore, exploit those determined to be reliable and discount the remainder. Other potential uses include the ability to detect misleading information on various Government forms (e.g., visa applications) that would suggest further investigation is warranted, to identify foreign sources who provide different information to home audiences and to the United States, and to identify false or misleading statements in textual documents.

In FY 2002, researchers under SBIR contracts demonstrated the ability to detect public corporations that might be potential targets of Securities and Exchange Commission (SEC) investigations, based on their SEC filings, well in advance of actual SEC investigations. They also demonstrated the ability to distinguish between news reports of deaths in a particular country as suicides or murders, depending on whether the sources were the official news agency or independent reports. In FY 2003, the MInDet Program will explore a number of techniques for detection of intentional misinformation in open sources, including linguistic genre analysis, learning with background knowledge, business process modeling, and adversarial plan recognition. In FY 2004, MInDet will select techniques with demonstrated ability to discriminate misinformation and transition them to selected intelligence and Defense users.

MINDET - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$3,000	\$5,000	\$12,000	FY 2007

PROGRAM SCHEDULE: MInDet begins in FY 2003 and concludes in FY 2007.

Milestone	FY/Quarter
Proof-of-Concept Studies	FY03 (2Q)
Proof of Concept Prototypes for Single Document Mis-Information Detection	FY03 (4Q)
Multiple Document Mis-Information Detection	FY04 (4Q)
Multiple Channel Mis-Information Detection	FY05 (3Q)
Multiple Author Mis-Information Detection	FY06 (4Q)
Multiple Language Mis-Information Detection	FY07 (4Q)

Human Identification at a Distance (HumanID) Program

The HumanID Program seeks to develop automated, multimodal biometric technologies with the capability to detect, recognize, and identify humans at a distance. DARPA believes that automated biometric recognition technologies could provide critical early warning support against terrorist, criminal, and other human-based threats. Obtaining this information may prevent or decrease the success rate of such attacks and provide more secure force protection of DoD operational facilities and installations. The HumanID Program seeks to develop a variety of individual biometric identification technologies capable of identifying humans at great distances in DoD operational environments. Once these individual technologies are developed, HumanID will develop methods for fusing these technologies into an advanced human identification system. This system will be capable of multimodal fusion using different biometric techniques with a focus on body parts identification, face identification, and human kinematics. Biometric signatures will be acquired from various collection sensors including video, infrared and multispectral sensors. These sensors will be networked to allow for complete

coverage of large facilities. The goal of this program is to identify humans as unique individuals (not necessarily by name) at a distance, at any time of the day or night, during all weather conditions, with noncooperative subjects, possibly disguised and alone or in groups.

HUMANID - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$16,710	\$11,120	\$4,325	\$000	FY 2004

PROGRAM SCHEDULE: The HumanID Program began in FY 2000 and will conclude in FY 2004.

Milestone	FY/Quarter
Initial development	FY01 (1-3Q)
In-situ evaluations	FY02 (1Q) FY02 (3Q)
Database development assessments	FY01 (2Q) FY02 (1Q) FY02 (3Q)
Biometric component evaluation	FY02 (1Q)
Decision milestone	FY02 (2Q)
Initial fusion experiments	FY03 (1Q)
Fusion experiments	FY04 (1Q)
Final technology evaluation	FY04 (1Q)

Activity, Recognition and Monitoring (ARM)

The ARM Program seeks to develop an automated capability to reliably capture, identify and classify human activities in surveillance environments. Currently, these types of activities are identified and analyzed by humans studying real-time and recorded video sequences. DARPA’s premise is that the capability to automatically identify and classify anomalous or suspicious activities will greatly enhance national security initiatives by providing increased warning for terrorist attacks, and increase the reconnaissance and surveillance capabilities for Intelligence and Special Operations Forces. ARM capabilities will be based on human activity models. From human activity models, the ARM Program will develop scenario-specific models that will enable operatives to differentiate among normal activities in a given area or situation and activities that should be considered suspicious. The program aims to develop technologies to analyze, model, and understand human movements, individual behavior in a scene, and crowd behavior. The approach will be multisensor and include video, agile sensors, low power radar, infrared, and radio frequency tags. The ARM Program will produce component technologies, and protosystems for demonstrating and evaluating performance for multiple scenarios. ARM is

a new program for FY 2004 that begins with new research areas identified in the HumanID Program.

ARM - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$000	\$5,500	\$9,500	FY 2008

PROGRAM SCHEDULE: The ARM Program begins in FY 2004 and concludes in FY 2008. A milestone schedule is under consideration.

Next-Generation Face Recognition (NGFR)

Face recognition technology has matured over the last decade, with commercial systems recognizing faces from frontal still imagery (e.g., mug shots). These systems operate in structured scenarios where physical and environmental characteristics are known and controlled. Performance under these conditions was documented in the Face Recognition Vendor Test (FRVT) 2000 and FRVT 2002. These evaluations demonstrated the advances in this technology; however, they also identified performance shortfalls in critical operational scenarios, including unstructured outdoor environments. The ability to operate in these operational scenarios is critical if these technologies are to be deployed in military, force protection, intelligence, and national security applications. DARPA believes that new techniques have emerged that have the potential to significantly improve face recognition capabilities in unstructured environments. These include three-dimensional imagery and processing techniques, expression analysis, use of temporal information inherent in video, and face recognition from infrared and multispectral imagery. The NGFR Program seeks to initiate development of a new generation of facially based biometrics that can be successfully employed in a wide variety of unstructured military and intelligence scenarios.

The major components of this program are a systematic development and evaluation of new approaches to face recognition; maturing of prototype systems at operational sites; experimentation on databases of at least one million individuals; and collection of a large database of facial imagery, which includes the variations in facial imagery found in unstructured environments. The NGFR Program aims to produce face recognition systems that are robust to time differences among facial imagery (aging) and variations in pose, illumination, and expression. NGFR is a new program for FY 2004 that begins with new research areas identified in the HumanID Program.

NGFR - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$000	\$7,000	\$10,140	FY 2007

PROGRAM SCHEDULE: The NGFR Program begins in FY 2004 and concludes in FY 2007. A milestone schedule is under consideration.

TIA Efficacy

The Promise of TIA

The Terrorism Information Awareness effort is an R&D program focused on a system/network concept. DoD's efforts are premised on the notion that individual and collective performance of those dealing with the terrorist threat can be improved dramatically with the assistance of computer tools working in a system/network environment.

The counterterrorism problem is characterized by new challenges for intelligence analysts, operators, and policy makers. More than ever before, attempts to "connect the dots" quickly overwhelm unassisted human abilities. The potentially important data sets are massive. The patterns sought are sparse, yet they may be anywhere in huge temporal and spatial regions. Frequently, analysts do not know what they are looking for.

DARPA believes that current stovepipe systems do not allow appropriate analysts to have access to all relevant information. Human limitations, biases, and other frailties often lead to consideration of a small part of the data that is available, failure to fully enumerate and evaluate the range of possibilities and outcomes, and failure to provide for adequate consideration of different points of view. The net result can be devastating.

In sum, neither individuals nor teams of unaided humans can function with maximum effectiveness in the present environment.

DARPA's aim in TIA research and development is to seek a revolutionary leap forward by augmenting human performance in dealing with several facets of the terrorist problem. Through an aggressive program to harness and integrate a group of computer tools in various stages of R&D, DARPA plans to assist humans cope with massive and varied data sets, think and reason about the counterterrorism problem, and work together in ad hoc teams to bring diverse points of view to the solutions of the problems. By augmenting human performance using these computer tools, the TIA Program expects to diminish the amount of time humans must spend in discovering information and allow humans more time to focus their powerful intellects on things humans do best—thinking and analysis.

If successful, the TIA research and development effort will demonstrate that some or all the tools under development really do contribute to the successful accomplishment of the counterterrorism mission—in particular, dramatically improve the predictive assessments of the plans, intentions, or capabilities of terrorists or terrorist groups. If successful, TIA and its component tools would foster the following five goals:

- **Secure Collaborative Problem Solving:** Would enable ad hoc groups to form quickly within and across agency boundaries to bring relevant data, diverse points of view, and experience to bear in solving the complex problems associated with countering terrorism.

- **Structured Discovery with Sources and Methods Security:** Would aid in the process of discovering planning and preparation for international terrorist attacks against the United States at home and abroad by examining transactions that may be made in carrying out these planning and preparation activities. If appropriate and lawful, DARPA envisions that large data sources including open source and classified intelligence information could be examined under appropriate strictures, rules, and oversight mechanisms.
- **Link and Group Understanding:** Would help identify terrorists and terrorist groups by discovering linkages amongst people, places, things and events related to suspected terrorist activities.
- **Context Aware Visualization:** Would make the information more understandable in a shorter time and by viewing data in new ways would help reveal otherwise undetected information such as patterns of activities that may be detected only by an experienced analyst.
- **Decision Making with Corporate Memory:** Would deliver to the decision-maker an understanding of history as well as an understanding in breadth and depth of the plausible outcomes of the current situation including a risk analysis of the various actionable options.

How TIA Would Work

For an understanding of the potential benefits that DoD believes may be achieved with TIA, it is important to understand how DoD envisions it would work if implemented. Teams of very experienced analysts and other experts (a red team) would imagine the types of terrorist attacks that might be carried out against the United States at home or abroad. They would develop scenarios for these attacks and determine what kind of planning and preparation activities would have to be carried out in order to conduct these attacks. These scenarios (models) would be based on historical examples, estimated capabilities, and imagination about how these tactics might be adapted to take into account preventive measures the United States has in place. The red team would determine the types of transactions that would have to be carried out to perform these activities. Examples of these transactions are the purchase of airlines tickets for travel to potential attack sites for reconnaissance purposes, payment for some kind of specialized training, or the purchase of materials for a bomb. These transactions would form a pattern that may be discernable in certain databases to which the U.S Government would have lawful access. Specific patterns would be identified that are related to potential terrorist planning. It is not a matter of looking for unusual patterns, but instead searching for patterns that are related to predicted terrorist activities.

Analysts from the Intelligence Community would use these models and other intelligence to guide their use of discovery tools to search, as appropriate, the permitted databases available to their respective communities. Procedures and techniques would be in place to protect the security of sensitive intelligence sources and, where applicable, the anonymity of U.S. persons if

access to these types of databases were ever contemplated. The databases may contain various forms of data including video, text, and voice in foreign languages. Relevant data would be transcribed and translated into English.

The analysts would work together using computer tools that allow them to remain with their parent organizations, yet meet in virtual spaces (something like an Internet chat room) to reason about a particular problem and share ideas and information related to the problem.

Other computer tools would identify linkages and relationships with other potentially relevant information. Requirements for collecting specific new intelligence to verify or refute the hypothesis being developed would be identified. There will always be uncertainty and ambiguity in interpreting the information available. Thus, different hypotheses would be developed by the analysts to reflect their differing points of view. These “competing hypotheses” would be passed to other groups of analysts working in similar virtual spaces in the operations and policy communities where they would estimate what these hypotheses might mean for a range of plausible future attacks. Options for taking actions to prevent the broadest range of plausible attacks would be developed. Analyses to determine the risks involved in taking these actions would be developed. Computer tools would assist the analysts in reasoning about all these issues and preparing the case for the decision-maker. Finally, all this information would be presented to the decision-maker in a manner and form that makes it quickly and easily understood even though these are almost always complex issues.

The overall objective would be to get the facts and issues before the decision-maker as early as possible so the decision-maker has the maximum number of viable options. TIA and its supporting programs are working on computer tools to aid the humans in all stages of this process. No stage of the analysis would stand by itself.

Measuring TIA Progress and Effectiveness

Funding for TIA research and development began in FY 2003. It is very early in the prototype TIA system/network development process to fully assess its efficacy; however, detailed plans are in place to evaluate the added value of a TIA-like system/network if it were made fully operational. As the R&D and experiments continue, DoD will establish quantitative measures of this added value. This is a fundamental purpose of R&D. Some anecdotal views have been captured during the limited experiments conducted to date.

The major problem in measuring added value in a system/network such as TIA is we seldom know the actual truth of the situation. We can never know for certain that there is a terrorist plan out there to be detected until after the fact; therefore, DoD is developing collateral measures of performance. The TIA R&D plan to measure added value is divided into four categories. DARPA is developing measures that help it understand performance in these categories.

- Technical. Processing-related system goals; e.g., numbers of documents ingested, patterns discovered, associations identified, and data sources investigated.

- Operational. How the technologies enhance the ways analysts approach their missions.
- Cognitive. How a technology can effectively increase an analyst's time for thinking as well as the true effect of a technology in this environment by normalizing and validating anecdotal evidence that demonstrates how the computer tools assist the analysts in accomplishing their missions more effectively.
- Network Interactions. Different ways analysis teams use the network to work together.

Researchers will assess the value of individual metrics within these categories in focused experiments. These metrics measurements were started in December 2002, and are just becoming established. This evaluation process will help guide the R&D and eventually influence implementation decisions.

The infrastructure and collection of software tools to be tested and evaluated under the 5-year R&D program are at varying levels of maturity. Some tools are ready for preliminary testing and evaluation, while others will require considerable R&D. At the beginning of the TIA Program, authorization was obtained to establish a virtual private network (VPN) over one of the classified DoD operational networks. The authorization included the ability to use experimental software on this VPN. Agreement has been reached with nine agencies and commands of the intelligence, counterintelligence, and military operational communities to participate in this experimental network. (These entities are listed on page 17.) The tools from the supporting programs that were ready for testing and evaluation were installed. These tools were supplemented with some from commercial sources. The most significant objective achieved is the establishment of a collaborative environment in which these participants can form ad hoc groups across the organizations, discover new experts and ideas, and begin to work operational problems in the global war on terrorism such as:

- Analyzing data from detainees from Afghanistan and finding relationships among entities in that data and with additional relationships from all-source foreign intelligence information.
- Assessing various intelligence aspects including weapons of mass destruction in the Iraqi situation.
- Aggregating very large quantities of information based on patterns into a visual representation of very complex relationships, which enabled rapid discovery of previously unknown relationships of operational significance.

The introduction of a systematic way of addressing these problems through structured argumentation has enabled a rapid understanding of issues and engendered prompt input from the various organizations. One organization may not have all the expertise required to address issues, but can quickly obtain assistance from others who do have the expertise.

The introduction of easy-to-use collaboration tools has slowly begun to change the way analysts find expertise to help them answer a question or resolve a discrepancy. They are becoming less hesitant to reach out to other acknowledged experts and participate in online discussions of the issues. Documents and pointers are provided. The result is a deeper understanding and a measurable increase of the supporting evidence for a position—all gained in reduced time.

The collaboration tools are also facilitating the rapid use of feedback from the results of higher-level analyses to adjust the filter parameters used on the incoming data.

Experiments have focused on automatically filtering very large amounts of foreign intelligence data to find relevant information in order to reduce the amount of material that must be read by analysts. DoD believes that the results of these initial experiments are very impressive and have revealed information that was not otherwise detected. The details of these experiments are classified and are available in a classified briefing.

The most significant measure of future potential is the interest and participation of the nine organizations of the experimental network.

- U.S. Army Intelligence and Security Command (INSCOM)
- National Security Agency (NSA)
- Defense Intelligence Agency (DIA JITF-CT)
- Central Intelligence Agency (CIA)
- DoD's Counterintelligence Field Activity (CIFA)
- U.S. Strategic Command (STRATCOM)
- Special Operations Command (SOCOM)
- Joint Forces Command (JFCOM)
- Joint Warfare Analysis Center (JWAC)

These represent a critical cross section of the relevant user domains that are involved in counter-terrorism.

Status of Component Research

The development, testing, and evaluation of some computer tools are in very preliminary stages and are being conducted in the individual component programs rather than in TIA. Some of this testing involves technologies to find specific patterns of transactions that are related to terrorist planning activities. In these cases, testing involves the use of synthetic data by research entities rather than real data by operational users. A portion of this research is addressing the problems of false alarms. DARPA is faced with a very difficult problem and only through research will DARPA be able to determine whether it is possible to find these sparse pieces of evidence in the vast amount of information about transactions with an accuracy that can be managed successfully in later stages of analysis. DARPA is just beginning these tests and does not yet have any results to report.

Laws and Regulations Governing Federal Government Information Collection

Public Law 108-7 requires that this report “set[] forth a list of the laws and regulations that govern the information to be collected by the Total Information Awareness program.”

If and when the TIA Program succeeds in developing technologies that operational agencies may wish to deploy in the effort to detect and preempt terrorist activity, those agencies may need to retrieve specific information from a variety of sources, including, for example, records of transactions such as airline reservations. In addition to the restrictions imposed by various provisions of the Constitution of the United States, such as the Fourth and Fifth Amendments, there are numerous statutory, regulatory, and other legal constraints upon the accessing or gathering of information by Federal Agencies. While few, if any, statutes flatly prohibit government access to information, Congress has often prescribed particularized procedures for obtaining information that falls within specific categories.

We interpret Congress’s mandate to set forth “a list of the laws and regulations that govern the information to be collected by” the TIA Program to be a directive to enumerate the statutes and regulations that would constrain any future data collection by federal agencies if and when they began to deploy the information technology the TIA program had developed. To the extent that this list goes beyond the requirements of Public Law 108-7, we have erred on the side of being over-inclusive.

This task has been accomplished substantively by the Congressional Research Service (CRS) of the Library of Congress, in its *Report for Congress: Privacy: Total Information Awareness Programs and Related Information Access, Collection, and Protection Laws* (updated version March 21, 2003) (the “*CRS Report*”). The *CRS Report* states (at CRS-5), and we agree, that

“. . . federal law tends to employ a sectoral approach to the regulation of personal information. . . . These laws generally carve out exceptions for the disclosure of personally identifiable information to law enforcement officials and authorize access to personal information through use of search warrants, subpoenas, and court orders. Notice requirements vary according to statute.”

The *CRS Report* identifies and summarizes at some length a large number of Federal statutes that regulate access to personal information. See *CRS Report* at CRS-6—16; CRS-21—29. The statutes identified by the CRS comprise those that are likely to have the most significant impact on any future deployment by the operational agencies of technology developed by the TIA Program. In addition to the laws noted in the *CRS Report*, we have identified, and summarize below, further statutory and regulatory provisions that constrain certain types of data collection by Federal Agencies. In doing so, *we do not in any way suggest that TIA’s search tools should be authorized to analyze all these forms of data*; quite the opposite is true. Our point—and what we understand Congress to have intended for us to do—is to enumerate the laws that protect various kinds of information and that might either constrain or (as a logistical matter) completely block deployment of TIA search tools with respect to such data.

The Fourth Amendment of the United States Constitution imposes fundamental limits on the types of searches and seizures that may be conducted, and the Fifth Amendment requires that due process of law be afforded. In addition, the following statutes, all identified and described in general detail in the *CRS Report*, may be listed:

- Privacy Act, 5 U.S.C. § 552a, as amended by the Computer Matching and Privacy Protection Act of 1988, 5 U.S.C.A. § 552a note
- Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g
- Cable Communications Policy Act of 1984, 47 U.S.C. § 551
- Video Privacy Protection Act of 1988, 18 U.S.C. § 2710
- Telecommunications Act of 1996, 47 U.S.C. § 222
- Health Insurance Portability and Accountability Act of 1996, 42 U.S.C. § 1320d, *et seq.*, together with the Department of Health and Human Service's implementing regulation, *Standards for Privacy of Individually Identifiable Health Information*, 45 C.F.R. Pts. 160, 164
- Driver's Privacy Protection Act of 1994, 18 U.S.C. § 2721
- Title III of the Omnibus Crime Control and Safe Streets Act of 1968, as amended by the Electronic Communications Privacy Act of 1986, 18 U.S.C. §§ 2510 *et seq.*
- Foreign Intelligence Surveillance Act of 1978, 50 U.S.C. §§ 1861 *et seq.*
- Electronic Communications Privacy Act of 1986, 18 U.S.C. §§ 2701 *et seq.*
- Pen Registers and Trap and Trace Devices Act, 18 U.S.C. § 3121 *et seq.*
- U.S.A Patriot Act of 2001, Pub. L. No. 107-56
- Homeland Security Act of 2002, Pub. L. No. 107-296
- Fair Credit Reporting Act of 1970, 15 U.S.C. §§ 1681 *et seq.*
- Right to Financial Privacy Act of 1978, 12 U.S.C. §§ 3401 *et seq.*
- Gramm-Leach-Bliley Act of 1999, 15 U.S.C. §§ 6801 *et seq.*
- Children's Online Privacy Protection Act of 1998, 15 U.S.C. § 6501

The *CRS Report* further notes that a variety of category-specific statutes regulate the use and disclosure of particular types of information held by the Federal Government, such as restrictions on the disclosure of tax returns, 26 U.S.C. § 6103, or on information collected by the Census Bureau, 13 U.S.C. § 221.

In addition, we note the following statutes, regulations, and other materials. We do not intend to suggest that authorization be given to use TIA's search tools with respect to such data; our point is to enumerate the major statutes protecting many particularly sensitive types of information (statutes that, in many cases, might effectively prevent the use of TIA search tools).

STATUTES:

- Child Victims' and Child Witnesses' Rights (18 U.S.C. § 3509): In cases where a child (a person under the age of 18) is or is alleged to be a victim of a crime of physical abuse, sexual abuse, or exploitation, or is a witness to a crime committed against another person, all documents that disclose the name or any other information concerning a child must be kept in a secure place to which no person who does not have reason to know their contents has access. Further, these documents or the information in them that concerns a child can be disclosed only to persons, who, by reason of their participation in the proceeding, have reason to know such information. These restrictions apply to law enforcement personnel as well, including employees of the Department of Justice (DOJ). The name or other information concerning a child may be disclosed to the defendant, the attorney for the defendant, a multidisciplinary child abuse team, a guardian *ad litem*, or an adult attendant, or to anyone to whom, in the opinion of the court, disclosure is necessary to the welfare and well-being of the child.
- Federal Juvenile Delinquency Act (18 U.S.C. §§ 5031 *et seq.*): The Federal Juvenile Delinquency Act contains a provision at § 5038 which limits the release of records compiled during federal juvenile delinquency proceedings. The records may only be released (and only to the extent necessary) to respond to: (1) inquiries from another court, (2) inquiries from an agency that is preparing a presentence report for another court, (3) inquiries from law enforcement agencies if the request is related to a criminal investigation or to employment in that agency, (4) inquiries from the director of a treatment or detention facility to which the juvenile has been committed, (5) inquiries from an agency considering an applicant for a national security position, and (6) inquiries from the victim or the deceased victim's family about the disposition of the juvenile by the court.
- Acquisition, Preservation, and Exchange of Identification Records and Information (28 U.S.C. § 534): This Act requires the Attorney General to acquire, collect, classify, and preserve identification, criminal identification, crime, and other records and exchange such records and information with and for the official use of, authorized officials of the Federal Government, the States, cities, and penal and other institutions. The exchange of records and information is subject to cancellation if dissemination is made outside the receiving departments or related agencies.

- Financial Crimes Enforcement Network (31 U.S.C. § 310): This Act establishes the Financial Crimes Enforcement Network (FinCEN) as a bureau in the Treasury Department. It authorizes FinCEN to maintain a government-wide data access service to several categories of privately and publicly maintained financial information and to records and data maintained in Federal, state, local, and foreign governmental agencies, including information regarding national and international currency flows. FinCEN is to analyze and disseminate the available data in accordance with applicable legal requirements and Treasury Department guidelines in order to identify possible criminal activity, support ongoing investigations, prosecutions, and other proceedings, support intelligence or counterintelligence activities to protect against international terrorism, and for other purposes. Treasury Department operating procedures in accordance with the Privacy Act and the Right to Financial Privacy Act of 1978 are to establish standards and guidelines for determining who is to be given access to FinCEN data and what limits are to be imposed on the use of such information, and for screening out of the data maintenance system information about activities or relationships that involve or are closely associated with the exercise of constitutional rights.
- Alcohol and Drug Abuse Records (42 U.S.C. § 290dd-2) and Drug Test Results (Pub. L. No. 100-71, § 503): The Title 42 provision mandates that certain alcohol and drug abuse patient records may be disclosed, absent consent, only under certain circumstances: (1) to medical personnel in a bona fide emergency; (2) to qualified personnel for scientific research (but personnel may not directly or indirectly identify an individual patient in a report of such research); or (3) under order of a court of competent jurisdiction. Section 503 mandates that the results of a drug test of a Federal employee may be disclosed, absent consent, only under certain circumstances: (1) to the employee's medical review official; (2) to the administrator of any employee assistance program in which the employee is receiving counseling or treatment or is otherwise participating; (3) to any supervisory or management official within the employee's agency having authority to take adverse personnel action against such employee; or (4) pursuant to the order of a court of competent jurisdiction where required by the U.S. Government to defend against any challenge against any adverse personnel action.
- Americans with Disabilities Act and the Rehabilitation Act (42 U.S.C. §§ 12111-12117; 29 U.S.C. §§ 701-797; 38 U.S.C. §§ 2011-2014; 5 U.S.C. § 2301, § 2302; Exec. Order No. 11478, as amended by Exec. Order No. 12106): Under applicable Federal law, the improper release of medical information, whether inside or outside an agency, may be considered an act of disability discrimination.² Several Federal laws prohibit employment discrimination against disabled employees or job applicants because of their disabilities: (1) the Americans with Disabilities Act of 1990 (ADA) which applies, in general, to private and state and local government

² Although the Federal Government is excluded from the definition of "employers" covered by the ADA, the standards of Title I of that Act still apply to Federal employers through the Rehabilitation Act. Federal Agencies are prohibited from discriminating based on physical or mental disability by Section 501 of the Rehabilitation Act. The standards for determining whether Section 501 has been violated are the same as those applicable to the ADA.

employers; (2) the Rehabilitation Act of 1973, which applies to Federal contractors, private employers receiving Federal funds, and the Federal Government; (3) the Vietnam-Era Veterans Readjustment Assistance Act, which applies to federal contractors and subcontractors and the Federal Government; and (4) the Federal civil service statutes and related Executive Orders.

- The National Security Act of 1947: The National Security Act contains a number of provisions that affect the ability of Federal law enforcement agencies to share information.
 - 50 U.S.C. § 435: This statutory provision directs the President to establish procedures to govern access to classified information. The Act requires that these procedures limit access to those Executive Branch employees who have cleared an appropriate background investigation. These procedures were established by Executive Order 12958, signed on April 17, 1995; that Order was comprehensively amended by Executive Order 13292, signed March 25, 2003. Both Orders are discussed below.
 - 50 U.S.C. § 403-3(c)(6): This statutory provision gives the Director of Central Intelligence (DCI) the responsibility for “protect[ing] intelligence sources and methods from unauthorized disclosure.” The DCI exercises this authority by issuing “Director of Central Intelligence Directives” (DCIDs) that address security procedures, protection of information, etc. The DCIDs also apply to the intelligence elements of the Federal Bureau of Investigation (FBI) and the handling of classified information within the FBI generally.
 - 50 U.S.C. § 403(g): This statutory provision details the responsibilities of the Assistant Director of Central Intelligence for Analysis and Production. These responsibilities, among others, include oversight of the analysis and production of intelligence by the Intelligence Community; establishing standards and priorities; and monitoring the allocation of resources for analysis and production within the Intelligence Community.
 - 50 U.S.C. § 421: This statutory provision criminalizes the identification of a covert agent to any unauthorized individual.

FEDERAL RULES OF PROCEDURE:

- Federal Rule of Criminal Procedure 6(e): Rule 6(e) prohibits government attorneys who supervise grand juries from disclosing “matters occurring before the grand jury,” except under the limited circumstances enumerated in the Rule itself. Law enforcement personnel may gain access to grand jury material under the exception to secrecy set forth in Rule 6(e)(3)(A)(ii), which allows disclosure otherwise prohibited to be made to government personnel deemed necessary by an attorney for the government to assist that attorney in the performance of his/her duty to enforce federal criminal law.

Section 203 of the U.S.A Patriot Act amended Rule 6(e) to permit the disclosure of grand jury information involving intelligence information “to any Federal law

enforcement, intelligence, protective, immigration, national defense, or national security official in order to assist the official receiving that information in the performance of his official duties.” This section requires subsequent notice to the court of the agencies to which information was disseminated and adds a definition of “foreign intelligence information” to Rule 6(e). This section also requires the Attorney General to develop procedures for the sharing of grand jury information that identified a U.S. citizen. The Attorney General issued the required *Guidelines for Disclosure of Grand Jury and Electronic, Wire, and Oral Interception Information Identifying United States Persons* on September 23, 2002. The provision dealing with the sharing of grand jury information (§ 203(a)) is not subject to the sunset provision of the Patriot Act.

- Federal Rule of Criminal Procedure 32: A probation officer must prepare a presentence report and present it to the court before a sentence is imposed. The report includes such information as the defendant’s criminal history, financial condition, and a recommended sentencing range. The report is furnished to the defendant, his/her attorney, and the attorney for the Government for objections. The report cannot be submitted to the court or its contents disclosed to anyone unless the defendant has consented in writing, has pleaded guilty or *nolo contendere*, or has been found guilty.

EXECUTIVE ORDERS:

- Executive Order 12333: This Order governs the conduct of intelligence activities, including intelligence analysis, to provide the President and the National Security Council with the necessary information to develop foreign, defense, and economic policy to protect U.S. interests from foreign security threats. It seeks to protect the rights of U.S. persons. It requires the Director of Central Intelligence (DCI) to ensure the establishment by the Intelligence Community of common security and access systems for managing and handling foreign intelligence systems, information, and products; to ensure the timely exploitation and dissemination of data gathered by national foreign intelligence collection means; and, in accordance with law and relevant procedures approved by the Attorney General, to give the heads of the departments and agencies access to all intelligence developed by the Central Intelligence Agency (CIA) or staff elements of the DCI relevant to the national intelligence needs of the departments and agencies. Other departments and agencies, including the State Department, Treasury Department, DoD, and FBI are tasked with specific information collection and dissemination functions.

The Order further authorizes agencies within the Intelligence Community to collect, retain, or disseminate information concerning U.S. persons only in accordance with procedures approved by the Attorney General. Information of several kinds relating to U.S. persons may be collected, retained, and disseminated, including information that is publicly available; information constituting foreign intelligence or counterintelligence; information obtained in the course of a lawful foreign intelligence, counterintelligence, or international terrorism investigation; information needed to protect foreign intelligence or counterintelligence sources or methods from unauthorized disclosure; and information arising out of a lawful personnel, or

physical or communications security investigation. Intelligence Community agencies are directed to use the least intrusive collection techniques feasible within the United States or against U.S. persons abroad. Certain information collection techniques may not be used except in accordance with procedures approved by the Attorney General; other particular techniques are not permissible.

- Executive Orders 12958 and 13292: These Orders, referenced above, create an orderly system for handling classified information. Information is classified based on the damage that unauthorized disclosure would cause to national security, which includes defense against transnational terrorism. The most sensitive information is restricted to the smallest group of people with a need to know. The classification level of information is controlled by the agency that owns the information. The “third agency rule” provides that an agency receiving classified information must obtain the approval of the disseminating agency prior to any further dissemination. Further safeguards to restrict access and prevent unauthorized access or disclosure are required. In particular circumstances, the Departments of State, Defense, and Energy and the CIA may establish special access programs. It is a crime to disclose certain classified information (pertaining to cryptographic or communication intelligence activities) to an unauthorized person. See 18 U.S.C. § 798.

REGULATIONS:

- 28 CFR 100.20 Confidentiality of Trade Secrets/Proprietary Information: Any company proprietary information provided to the FBI under this part shall be treated as privileged and confidential and shared only within the government on a need-to-know basis. It shall not be disclosed outside the government for any reason inclusive of the Freedom of Information requests, without the prior written approval of the company.

DEPARTMENT OF JUSTICE GUIDANCE / ORDERS:

- Attorney General Guidelines on General Crimes, Racketeering Enterprise and Domestic Security/Terrorism Investigations: These Guidelines were revised on May 30, 2002, and provide guidance for general crimes and criminal intelligence investigations by the FBI. The standards and requirements set forth therein govern the circumstances under which such investigations may begin and the permissible scope, duration, subject matters, methods, and objectives of these investigations.
- Attorney General Guidelines Applicable to FBI Foreign Counterintelligence Investigations: The FBI may disseminate information under these guidelines to other Federal agencies if the information relates to a crime or violation of law or regulation that falls within the recipient agency’s investigative jurisdiction, otherwise relates to the recipient agency’s authorized responsibilities, is required to be furnished by Executive Order 10450, or is required to be disseminated by statute, Presidential directive, National Security Council directive, or an interagency agreement that has been approved by the Attorney General. The FBI may disseminate information to

state and local governments with appropriate jurisdiction if such dissemination is consistent with national security. Dissemination to a foreign government is permitted under specified circumstances, as is dissemination to Congressional committees and the White House.

- Attorney General Guidelines Regarding Disclosure to the Director of Central Intelligence and Homeland Security Officials of Foreign Intelligence Acquired in the Course of a Criminal Investigation: These guidelines were issued on September 23, 2002, pursuant to § 905(a) of the U.S.A Patriot Act. The guidelines formalize a framework pursuant to § 905(a) for facilitating and increasing the expeditious sharing of foreign intelligence acquired in the course of criminal investigations.
- DOJ 1792.1B Chapter 4, Maintenance of Records and Reports Systems, Alcohol and Drug Abuse Records: The DOJ's policy is one of nondisclosure of client records, except to the extent that nonconsensual disclosure is authorized by law or to the extent necessary to prevent an imminent and potential crime which directly threatens loss of life or serious bodily injury.
- DOJ 1900.5A National Security Emergency Preparedness Program and Responsibilities: The FBI is responsible for providing a response to foreign counterintelligence and domestic security and terrorism threats that includes (1) disseminating information, to the extent that conditions permit, concerning hostile intentions and activities toward government officials and agencies and (2) responding to specific requests from senior government officials and agencies for FBI information related to foreign counterintelligence and domestic security matters.
- DOJ 2620.5A Safeguarding Tax Returns and Tax Return Information: Employees of the DOJ to whom tax return information is entrusted are responsible for its safeguarding and are prohibited from disclosing such information except as permitted by law. Tax information shall not be disseminated to, discussed with, or exposed to unauthorized persons.
- DOJ 2620.7 Control and Protection of Limited Official Use Information, Dissemination and Transmission: Information which has been identified and is known by recipient as "Limited Official Use" shall be safeguarded from disclosure to unauthorized individuals whether or not the material is physically marked. Safeguarding from disclosure includes precautions against oral disclosure, prevention of visual access to the information, and precautions against release of the material to unauthorized personnel.
- DOJ 2640.1 Privacy Act Security Regulations for Systems of Records: This order applies to all DOJ organizations that maintain systems of personal records.

DEPARTMENT OF DEFENSE REGULATIONS AND GUIDANCE:

- DoD 5240.1-R Procedures Governing the Activities of DoD Intelligence Components That Affect United States Persons: These procedures, which were approved by the Attorney General, implement Executive Order 12333.
- DoD 5200.27 Acquisition of Information Concerning Persons and Organizations not Affiliated with the Department of Defense: This directive governs the acquisition of information by DoD components other than those with intelligence and counterintelligence responsibilities. DoD components are prohibited from collecting, reporting, processing, or storing information on individuals or organizations not affiliated with DoD, except when such information is essential to the accomplishment of specified DoD missions.
- 32 CFR 311, 312, 318, 319, 321 through 323, 326, 505, 701.100, and 806b, Exemption of Records under the Privacy Act: The referenced DoD systems of records are exempt from various requirements of the Privacy Act. Each Part of the CFR identifies a DoD Component, such as the Army, Defense Security Service, Defense Intelligence Agency, etc., which has claimed an exemption for the record system identified.
- 32 CFR 310, DoD Privacy Program: This regulation governs how the DoD protects records covered by the Privacy Act, and under what conditions, it may, absent consent of the individual about whom the records pertain, disclose such records.
- 32 CFR 275, Obtaining Information From Financial Institutions: This regulation governs the procedures for the DoD to use to gain access to financial records maintained by financial institutions.

TIA's Impact on Privacy and Civil Liberties, and Recommended Practices, Procedures, Regulations or Legislation for TIA Deployment and Implementation to Eliminate or Minimize Adverse Effects

Overview

Public Law 108-7 requires that this report “assess[] the likely impact of the implementation of a system such as the Total Information Awareness program on privacy and civil liberties.”

Preliminary to any such analysis, DoD wishes to make certain points clear. In seeking to develop innovative information technology that DoD hopes will improve the nation's capabilities to detect, deter, preempt, and counter terrorist threats, TIA's research and testing activities have depended *entirely* on (1) information legally obtainable and usable by the Federal Government under existing law, or (2) wholly synthetic, artificial data that has been generated to resemble and model real-world patterns of behavior. Further, the TIA Program is not attempting to create or access a centralized database that will store information gathered from various publicly or privately held databases.

Nevertheless, ultimate implementation of some of the component programs of TIA may raise significant and novel privacy and civil liberties policy issues. Largely because of the greater power and resolution of TIA's search and data analysis tools, questions will arise concerning whether the safeguards against unauthorized access and use are sufficiently rigorous, and whether the tools can or should be applied at all with respect to certain types of particularly sensitive information. In addition, privacy and civil liberties issues may arise because some would argue that the performance and promise of the tools might lead some U.S. Government agencies to consider increasing the extent of the collection and use of information already obtained under existing authorities.

The DoD has expressed its full commitment to planning, executing, and overseeing the TIA Program in a manner that is protective of privacy and civil liberties values. Safeguarding the privacy and the civil liberties of Americans is a bedrock principle. DoD intends to make it a pervasive element in the DoD management and oversight of the TIA Program. These two sets of interests—privacy and civil liberties—are complementary, yet distinct. Privacy relates primarily to the right of the individual person to freedom from various forms of governmental intrusion and unwanted exposure of sensitive information; while civil liberties relate primarily to the protection of the individual's constitutional rights to, among others, freedom of expression, freedom of the press and assembly, freedom of religion, interstate travel, equal protection, and due process of law. The DoD's TIA work addresses both privacy and civil liberties in three principal ways:

- In its TIA work, as in all of its missions, the DoD must fully comply with the laws and regulations governing intelligence collection, retention, and dissemination, and all other laws, procedures, and controls protecting the privacy and constitutional rights of U.S. persons.

- TIA is seeking to develop new technologies, including Genisys Privacy Protection, that will safeguard the privacy of U.S. persons by requiring, documenting, and auditing compliance with the applicable legal requirements and procedures.
- TIA’s research and testing activities are conducted using either real information that the Federal Government has already legally obtained under existing legal regimes, or synthetic, wholly artificial information generated in the laboratory about imaginary persons engaged in imaginary transactions—data that by definition does not implicate the privacy interests of U.S. persons.

In addition to these measures, the DoD intends, as an integral part of oversight of TIA, to continuously monitor and assess emerging potential privacy and civil liberties issues. Because TIA is still largely in the developmental stage, any effort to identify such issues is, of necessity, preliminary. Nonetheless, we believe that certain overall privacy policy issues can be identified, and we have made preliminary recommendations below with respect to those issues.

As TIA research efforts move forward, examination of these issues will require a detailed and rigorous understanding of the particular tool and data involved, their present and potential future contributions to the public safety and other national interests, their impact on privacy values, and the legal, policy, technological, and human engineering checks and balances that are already in place as well as additional checks and balances that may be imposed on the use of the particular tool and data. Addressing these issues will lead to a careful determination of the correct course of action after assessing these values and interests in light of our Nation’s commitment to security *and* privacy. These issues will be illuminated by the progress of TIA in developing and testing tools by lawful means and applying these tools against both synthetically generated and lawfully acquired data.

To accomplish this objective of ongoing and effective oversight and review, a senior representative of the DoD will chair an oversight board. This oversight board and the Secretary of Defense will receive advice on legal and policy issues, including privacy, posed by TIA from a Federal Advisory Committee composed of outside experts.

This report does not recommend any changes in statutory law, but instead contemplates that any deployment of TIA’s search tools may occur only to the extent that such a deployment is consistent with current law. Accordingly, the strictures of current law protecting certain categories and sources of information may well constrain or (as a logistical matter) completely preclude deployment of TIA search tools with respect to such data.

Relevant Information Privacy Principles

As with any intelligence activity, the use of TIA tools and technologies by operational agencies must be conducted in accordance with all relevant regulations, statutes, and constitutional principles. Moreover, the development of TIA tools and techniques by DARPA must comply with all applicable laws. Above and beyond these basic legal requirements, however, a proper consideration and resolution of the privacy policy issues that are raised by TIA is necessary.

A proper analysis of the privacy policy issues that would be raised by deployment of TIA should first begin with some articulation of the general privacy *principles* that should guide that analysis. In light of the unspeakable terrorist acts to which our country has been subjected and the further terrorist threats we may face in the future, there can be no question but that the government must devise ways to better enable it to detect such threats before they occur. The question is how to accomplish that in a manner that preserves, and even strengthens, our basic commitment to privacy and civil liberties.

In a sense, one simple idea captures both sides of the coin in the security versus privacy debate: “Knowledge is power.” The more information the government has, the more it can find out about terrorists’ plans and act to prevent them. On the other hand, the more information the government has about our citizens, the more opportunities there are that such information could be seriously misused. The goal of any sensible information privacy policy must be to help to ensure that activities relating to information collection, storage, sharing, and analysis do not threaten privacy and civil liberties.

Any attempt to articulate overall policy principles concerning information privacy will necessarily be somewhat generalized. The answer in any given case will depend upon the particular issue and the competing values at stake. Nonetheless, some general considerations can be described that can help to structure and guide the analysis of such issues:

The importance of identifying the nature and magnitude of the particular privacy interests implicated

- There are a variety of different privacy interests, and they are not all of the same magnitude. Saying that something presents “privacy concerns” should be the beginning of an analysis as to the nature and severity of those concerns, the strength of the countervailing interests, and whether and how the privacy concerns identified can be mitigated. The basic concept of informational privacy includes several key concerns, not all of which are of the same degree and character. Among the most important are the following concerns:
 - *Access to particularly sensitive information.* Certain kinds of information about a person (e.g., medical records and tax records) are particularly sensitive, because access to such information presents serious opportunities for abuse. Most such sensitive categories of information are already covered by detailed statutory and regulatory regimes.
 - *Access to aggregate individually identifiable information.* Even when individual items of data are not particularly sensitive, access to an aggregation of significant quantities of personal data on specific persons presents opportunities for misuse and for unwarranted intrusion into personal matters.
 - *Maintaining and storing individually identifiable information.* The storage by the government of individually identifiable information, precisely because of its permanence, increases the practical possibilities for misuse of the information.

- *Capacity for unauthorized access to individually identifiable information.* Any system for accessing or storing personal information must be secure against intruders and other unauthorized users, who may seek to use it for improper purposes.
- *Capacity for unauthorized use of particular investigatory tools.* Consideration must be given as to whether there is anything about the particular characteristics or usage of a given tool that itself creates additional possibilities for misuse by persons who have authorized access.
- *Accuracy of individually identifiable information.* If inaccurate information is publicly disseminated, that may harm reputational interests, and if it is used as a basis for important decisions affecting the individual, it will have additional and potentially significant adverse impacts.

The importance of practical, operational safeguards

- When it comes to analyzing privacy issues, “thou shalt not” is good, but “thou cannot” is better. Anyone who has ever worked to design a system to protect valuable information (such as a trade secret) appreciates the need for internal operational safeguards that reduce the *opportunities* for mischief. There is a need to have legally enforceable prohibitions against any mischief that nonetheless occurs, but additional internal operational safeguards are also necessary.

Consideration of the weight of competing values

- In light of the nature and magnitude of the particular privacy interests implicated, the available practical means for mitigating those concerns, and an assessment of the actual practical value of the tools in question for protecting against terrorist threats, an evaluation must then be made as to whether particular deployments of the technology can be carried out in a way that achieves those objectives without sacrificing privacy.

Preliminary Assessment of Privacy Implications of TIA and Pertinent Recommendations

Introduction

With these basic principles in mind, some preliminary observations can be made about the likely impact of the implementation of TIA on privacy and civil liberties, and some recommendations concerning the measures that may be warranted to eliminate or minimize adverse concerns on privacy and other civil liberties. Because TIA is still largely in the developmental stage, these observations are, of necessity, preliminary.

DoD, however, wishes to emphasize two fundamental points at the outset. First, DoD must pursue any technological breakthroughs in the various TIA programs, which are described in this report, in full compliance with existing law. Second, the Department of Defense, the Department of Justice, and the Central Intelligence Agency take very seriously the obligation to protect privacy and civil liberties. Accordingly, any deployment of TIA tools would occur only after careful analysis of the relevant policy issues and in accordance with the recommendations set forth below.

One measure of the importance DoD attaches to privacy and civil liberties issues is reflected in the fact that, in addition to the other measures undertaken by DoD in analyzing these issues, the Secretary of Defense has sought the guidance of outside experts. DoD has established a Federal Advisory Committee to advise the Secretary of Defense on the legal and policy issues, particularly those related to privacy, that are raised by the application of advanced technologies to be used in the war on terrorism, such as TIA. This advisory committee is expected to hold its first meeting in late May 2003.

Particular TIA Programs that Have Raised Privacy Concerns

The privacy concerns that have been raised with respect to TIA focus on the data search and pattern recognition tools that are being researched. Broadly speaking, the data search, pattern recognition, and privacy protection programs include eight different technologies: Genisys, Evidence Extraction and Link Discovery (EELD), Scalable Social Network Analysis (SSNA), MisInformation Detection (MInDet), Bio-Event Advanced Leading Indicator Recognition Technology (Bio-ALIRT), Human Identification at a Distance (HumanID) Program, Activity Recognition Monitoring (ARM), and Next-Generation Face Recognition (NGFR).

These eight programs do not all raise the same issues or the same level of concern. Bio-ALIRT relies on using aggregate statistical data or anonymized data that eliminates concerns about individually identifiable data. DARPA affirms that use and collection of data by Bio-ALIRT must be done in accordance with all applicable laws. The various tools for human identification at a distance (HumanID, ARM, and NGFR) would raise significant privacy issues, depending upon their efficacy and accuracy, the places and circumstances in which they were deployed, and whether they were used to analyze (or to justify longer retention of) stored surveillance tapes of public places. DoD is committed to ensuring that these issues receive careful analysis as these programs move forward, but they are not the programs that have given rise to the greatest level of concern (or that gave rise to this report).

The primary privacy concerns raised about TIA focus on the data search and analysis tools: Genisys, EELD, SSNA, and MInDet. The privacy concerns raised by TIA's search tools, of course, will depend significantly upon the types of information contained in the databases for which use of these tools is authorized, and upon the authorities, procedures, and safeguards that are established. At the present time, the only tools from this category that are being used in TIA network tests come from the EELD Program and they are being applied only with respect to foreign intelligence data.

As research on the tools progresses and additional deployments are considered, different concerns will be raised depending upon the types of information in the authorized databases. If, for example, a particular deployment permitted only querying of databases on non-U.S. persons, that would present less concern than would querying for information about foreigners in databases that also happen to contain information on U.S. persons, which in turn would raise less concern than would querying about U.S. persons directly. With this important reservation in mind, a number of general observations can be made about the likely privacy concerns and the possible methods for analyzing and resolving those concerns.

Privacy Issues that TIA Does *Not* Raise

In analyzing the privacy issues that are raised by these particular TIA programs, it is important to recognize what they do *not* do.

- Nothing in the TIA Program changes anything about the types of underlying information to which the government either does or does not have lawful access, nor does it change anything about the standards that must be satisfied for accessing particular types of data. TIA does not grant the government access to data that is currently legally unavailable to it. On the contrary, any deployment of TIA would have to operate within the confines imposed by current law. Accordingly, to the extent that access to certain particularly sensitive categories of information is restricted by law, the deployment of TIA search tools with respect to such data would comport with such standards, or (depending upon the nature of the legal restriction) in some cases might be logistically infeasible altogether.
- As conceived, TIA's search tools, if and when used by operational agencies, would leave the underlying data where it is, extracting only what is responsive to a specific and defined query, and not engaging in random searches. While this does not eliminate all privacy concerns, this feature of TIA is an important and, on balance, privacy-enhancing logistical limitation, because the practical risks for misuse of personal data would be increased if complete possession and control of the relevant data were assumed by the government.
- Just as TIA would leave the underlying data where it is, it would, in terms of the substance of such information, take the data as it finds it. That is, nothing in the implementation of TIA envisions that parties whose databases would be queried should begin collecting data that they do not already collect. This avoids a significant privacy concern that would otherwise be present.
- It follows as a corollary to the previous points that TIA does not, in and of itself, raise any particular concerns about the accuracy of individually identifiable information. On the contrary, TIA is conceived of as simply a tool for more efficiently inquiring about data in the hands of others, and in theory these inquiries currently could be made by more labor-intensive human efforts. Although (quite apart from TIA) various concerns have been raised about the quality and accuracy of databases that are in private hands, these general concerns would exist regardless of the method

chosen to query these databases and, thus, do not present a concern specific to TIA. Of course, to ensure the accuracy and utility of any information retrieved by TIA's search tools, consideration should be given, in implementation, to the quality of the databases to be queried.

Privacy Issues Raised by TIA and Recommendations for Addressing these Issues

The primary privacy issues raised by TIA are threefold:

- Aggregation of data
- Unauthorized access to TIA
- Unauthorized use of TIA

To the extent that TIA's search tools would ever be applied to data sources that contain information on U.S. persons, the privacy issues raised by these tools are significant ones that would require careful and serious examination. As a logistical matter, there is a "practical obscurity" inherent in the dispersal of scattered bits of personal data. TIA's search tools have the capacity to eliminate this practical obscurity and to provide a user with quick access to a wide range of information. The potential benefits of such a tool in identifying terrorist activity could be significant. On the other hand, the potential harm that could result from misuse of this effective aggregation of large quantities of data are obvious. Several factors need to be considered in evaluating TIA's suitability for deployment in particular contexts.

- The *efficacy and accuracy* of TIA's search tools must be stress-tested and demonstrated. The tools must be shown to be sufficiently precise and accurate; i.e., a search query results in *only* that information that is responsive to the query. TIA's tools must be demonstrated to be sufficiently precise so that, if only a limited query is legally authorized, the data retrieved remains within the strictures of the law and the query does not grant access to data that may not lawfully be accessed. DARPA has expressed its commitment to the necessary testing to ensure the technological accuracy of TIA's search tools.³ Moreover, the Secretary of Defense has established an oversight framework governing the R&D phases of this project. To ensure the R&D activities being pursued under the TIA Program continue to be conducted in accordance with all applicable laws, regulations, and policies, the Secretary of Defense established in February 2003 an internal oversight board to oversee and monitor the manner in which TIA tools are being developed and prepared for transition to real world use. This board, composed of senior DoD and Intelligence Community officials, will establish policies and procedures for testing of the TIA-developed tools. In addition, the board will examine the various tools in light of existing privacy protection laws and policies and recommend appropriate program modifications to DARPA.

³ This particular efficacy concern is distinct from, and in addition to, the basic question of whether the TIA tools can produce the positive value contemplated. As made clear elsewhere in this Report, if the tools developed in TIA "cannot extract terrorist signatures from a world of noise, even for simulated data, then there is no reason to proceed." See *infra* at Appendix A-11.

- This is a situation in which the need for *built-in operational safeguards* to reduce the opportunities for abuse are absolutely critical. DARPA is already researching whether and how it may be able to build in controls that, at an architectural level, would govern TIA's search tools. Among the controls being researched are automated audit trails to document who accessed the system and how it was used during the session; anonymization of sources of data and of the persons mentioned in the underlying data, so these data could not be revealed unless it is lawful and warranted; selective revelation of data, so additional permissions would need to be obtained in order to receive additional data; and rigorous access controls and permissioning techniques. TIA's ultimate suitability for particular purposes will depend heavily upon DARPA's success on these technological issues.
- It will be essential to ensure that *substantial security measures* are in place to protect these tools from unauthorized access by hackers or other intruders. Some of these measures must be built-in at the architectural level; others will involve the adoption of policies that prescribe who may have access, for what purposes, and in what manner.
- Any agency contemplating deploying TIA's search tools for use in particular contexts will be required to conduct a *pre-deployment legal review* of whether the contemplated deployment is consistent with all applicable laws, regulations, and policies. Some particular deployments, for example, might only be legally permissible if the tools developed had been shown, as a technological matter, to properly avoid retrieving data on U.S. persons, whether through anonymization techniques or otherwise. In this regard, it should be noted that the DoD General Counsel has directed each operational component within DoD that hosts TIA tools or technologies to prepare a substantive legal review that examines the relationship between that component and TIA and analyzes the legal issues raised by the underlying program to which the TIA tools will be applied. The General Counsel also has advised that all such relationships should be documented in a memorandum of agreement between TIA and the component to ensure that the relationship is clearly understood by all parties. These memoranda of agreement with non-DoD components will specify that a similar legal review be conducted by the non-DoD component.
- There will be a need for any user agency to adopt policies establishing *effective oversight* of the actual use and operation of the system before it is deployed in particular contexts. This will include periodic and spot auditing and testing of the system, periodic review of its operation, restrictions on access to the system, and prompt and effective procedures for detecting and correcting misuse of the system and for punishing the violators. There must be clear and effective accountability for misuse of the system.

An additional privacy issue is whether there is anything about the particular *technological architecture* of the TIA tools that implicates specific privacy concerns, i.e., issues over and above those inherent in the overall nature of the task the tool is performing. One such issue relates to the manner in which the TIA tools would achieve interoperability with the databases with which they interact. If, for example, this would require installation of government-developed software code onto privately owned databases, this will raise a potentially significant privacy concern. Analysis of this issue would require a consideration of a number of different factors, including the feasibility of alternative mechanisms and whether transparency could be achieved, without loss of security, by making publicly available the underlying software code installed.

Finally, the various tools for human identification at a distance (HumanID, ARM, and NGFR) may raise significant privacy issues if deployed in particular contexts. As an initial matter, any deployment of these tools in the United States would need to be reviewed in advance in order to ensure compliance with the strictures of the Fourth Amendment. *Cf. Kyllo v. United States*, 533 U.S. 27 (2001) (use of infrared technology can constitute a “search”). In addition, certain privacy policy issues would need to be considered. These issues primarily relate to the accuracy of these tools, the potential concerns about aggregation of data, and concerns about misuse. Resolution of these issues requires an evaluation of whether these tools can be shown to be accurate for their intended purposes, whether a particular location would be appropriate for their use, and whether they would be used to analyze (or to justify longer retention of) stored surveillance tapes of public places. These issues should receive careful analysis as these programs move forward.

In closing, DoD would like to underscore its realization that the successful development and the effective deployment and use of TIA tools may pose additional specific and currently unidentifiable privacy policy issues. DoD believes that the best way to navigate these issues consistent with our Nation’s most cherished values is to pursue the development of the most effective and most privacy-protecting tools possible and to address privacy and civil liberties issues squarely and continually as they arise, in specific factual contexts and in full partnership with other Executive Branch agencies and the Congress. DoD has expressed its commitment to the rule of law in this endeavor and views the protection of privacy and civil liberties as an integral and paramount goal in the development of counterterrorism technologies.

Appendix A – Detailed Description of TIA and High-Interest TIA-Related Programs

The target date for the deployment of each program is the completion date listed, unless identified differently in the descriptive paragraphs. Besides TIA, other TIA-related programs considered as high interest within the context of this report include:

- Genisys
- Genisys Privacy Protection
- Evidence Extraction and Link Discovery (EELD)
- Scalable Social Network Analysis (SSNA)
- MisInformation Detection (MInDet)
- Human Identification at a Distance (HumanID)
- Activity, Recognition and Monitoring (ARM)
- Next Generation Face Recognition (NGFR)

Terrorism Information Awareness (TIA)

OVERVIEW: TIA is a Defense Advanced Research Projects Agency (DARPA) research program that will integrate advanced collaborative and decision support tools; language translation; and data search, pattern recognition, and privacy protection technologies into an experimental prototype network focused on the problems of countering terrorism through better analysis. If successful and transitioned to operational uses, this program of programs would provide decision- and policy-makers with information and knowledge about terrorist planning and preparation activities that would aid in preventing future international terrorist attacks against the United States at home and abroad. If deployed, a TIA-like system/network could provide the Department of Defense (DoD) and Intelligence Community with tools and methods to solve many of the problems that have been identified in the aftermath of the attacks against the United States on September 11, 2001, and that call for improved analysis in our continuing war against terrorism. The report of the Congressional Joint SSCI-HPSCI Inquiry into the Events of 9/11/01⁴ concludes that the failure to identify the threat prior to the attacks of September 11, 2001, had less to do with the ability of authorities to gather information than with their inability to analyze, understand, share, and act on that information.

The major problems that TIA research and development aim to address include: the difficulties of sharing of data across agency boundaries; mistaking absence of evidence for evidence of absence; confusing unfamiliar with improbable; having too many unknown unknowns, generating a single hypothesis versus competing hypotheses; and better exploitation of all permitted and open source information. DARPA believes that, in most cases, these problems exist in part because of a lack of applied technology to aid the human assessment and analytic processes. In today's world, the amount of information that needs to be considered far exceeds the capacity of the unaided humans in the system. Adding more people is not necessarily the

⁴ Final Report of the Joint SSCI/HPSCI Inquiry into the Events of 9/11/01 dated Dec 10, 2002

solution. In DARPA's view, we need to provide a much more systematic methodological approach that automates many of the lower level functions that can be done by machines guided by the human users and gives the users more time for the higher level analysis functions which require the human's ability to think.

TIA is one of the research and development programs of DARPA's Information Awareness Office (IAO), which was established in January 2002. IAO was formed to bring together, under the leadership of one technical office director, several existing DARPA programs that were largely focused on R&D in various information technologies relevant to DoD's future capabilities in combating the asymmetric threat, and for imagining and creating some new programs that would be needed to fully address the technology needs for a complete prototype system/network to respond to the terrorist threat (one kind of asymmetric threat) in the wake of September 11. TIA is the system/ network-level integration program, while other IAO programs are designed to provide technologies and components needed by TIA. TIA will integrate these technologies and provide them to various organizations for experiments and will assess their utility in operationally relevant contexts.

The TIA research and development program began in FY 2003. Funding for FY 2003 through FY 2005 as proposed in the FY 2004 President's Budget submission is \$53,752,000. A number of organizations in the Intelligence Community have shown great interest in working with the TIA research and development effort to test and evaluate technologies. The organizations already participating or planning to participate in the near future in TIA's spiral development and experiments include:

- U.S. Army Intelligence and Security Command (INSCOM)
- National Security Agency (NSA)
- Defense Intelligence Agency (DIA JITF-CT)
- Central Intelligence Agency (CIA)
- DoD's Counterintelligence Field Activity (CIFA)
- U.S. Strategic Command (STRATCOM)
- Special Operations Command (SOCOM)
- Joint Forces Command (JFCOM)
- Joint Warfare Analysis Center (JWAC)

DARPA is providing these agencies and commands with system/network infrastructure and concepts; software analytical tools; installing this software; providing training on its use; observing experiments; evaluating the performance of the software; and collecting user comments on needed changes, modifications, and additions to the software. The operational agencies and commands are providing facilities and personnel to conduct these experiments and they are using data available to them in accordance with existing laws, regulations and policies applicable to each of them.

In the TIA research and development vision, four user domains must work together to comprehensively counter the terrorist threat: intelligence, counterintelligence, operations, and policy. Three of these domains are represented in the above list of agencies and commands

participating in experiments with TIA. It is envisioned that a national security policy organization will be added to the experiments.

To help realize the TIA vision, five major investigation threads are being pursued and are driving much of the experimental activity in the TIA Program: secure collaborative problem solving, structured discovery with security, link and group understanding, context aware visualization, and decision making with corporate memory.

- **Secure Collaborative Problem Solving.** The premise in this thread is that a collaborative environment is needed that enables ad hoc groups to quickly form within and across agency boundaries to bring relevant data, diverse points of view, and experience to bear in solving the complex problems associated with countering terrorism. There is always going to be uncertainty and ambiguity in the data available. There will be differing interpretations of the data. Competing hypotheses need to be developed and supported by models that lay out specifically the evidence, rationale, and logic supporting or not supporting some hypothesis. These hypotheses need to be considered in developing ranges of plausible outcomes, actionable options, and risks to aid the decision making process. If sensitive information is to be shared, this environment must be secure and constructed in such a way that various classification levels and need-to-know are managed in a sensible way that gets the relevant information to the right people in an expeditious manner. The system/network must provide for agility in assembling these ad hoc analysis groups and, at the same time, preserve the control functions of our agencies and commands.
- **Structured Discovery with Sources and Methods Security.** International terrorist organizations must plan and prepare for attacks against the United States at home and abroad, and their people must make transactions in carrying out these planning and preparation activities. Examples of transactions that may be of interest are activities such as telephone calls, travel arrangements, and the acquisition of critical materials to be used in their attacks. Data about these events may well be buried in an enormous amount of data about routine worldwide activity that has nothing to do with international terrorism. In addition, there is a wider range of intelligence data, both classified and open source, that must be searched to find relevant information for understanding the terrorist intent. To have any hope of making sense of this, DARPA believes that there must be a more structured and automated way of approaching this problem. This would also assist in developing strictures, rules, and oversight mechanisms. One cannot just randomly search the data for clues about suspicious behavior. As conceived by DARPA, terrorist attack scenarios would be developed that take into account what has historically happened as well as the best estimates of how the terrorist will adapt to our preventive measures. Models of these scenarios would be developed and refined to identify what specific transactions they would have to make to carry out their attacks. These models would identify specific data and patterns for which secure, focused data search and discovery might be done. Because of the volume of data that may need to be sorted through quickly and accurately, automated structured discovery methods would be developed. Data comes in many different forms and languages. Voice data would be automatically

transcribed to text to make it more easily searchable by machines. Foreign languages would be automatically translated into English. Unstructured text would be given some structure by identifying and extracting entities such as the names of people, places, things and events buried in the text so machines may process the volumes of text. To make sensitive data more widely shareable and available, security methods to protect the integrity of sensitive intelligence sources and methods as well as privacy would be developed. Structured discovery is only the early stages of the analysis problem. Several later stages of analysis would be required to eliminate false leads, to refine the search and discovery process, and to establish a better understanding of terrorist intent.

- **Link and Group Understanding.** One of the characteristics of the terrorist threat is that their organizational structures are not well understood and are purposefully designed to conceal their connections and relationships. DARPA's premise is that by discovering linkages among people, places, things and events and by training software algorithms to recognize patterns of relationships that are representative of terrorist groups, it can help identify terrorists and terrorist groups with software tools that will contribute to understanding potential terrorist intent, methods of operation, and organizational dynamics. This process is much easier if there is a suspect as a starting point; however, as the terrorists adapt to preventive measures, there is increased likelihood of "sleeper" cells for which there are no known connections to known terrorists. Thus, DARPA aims to develop techniques for detecting patterns that are based on known or estimated terrorist planning and preparation activities. Some of the prototype tools, which are applicable in these situations, are being developed in one of the IAO programs and early versions have been used by INSCOM analysts to help analyze captured data from Afghanistan and elsewhere. Although this work is in the early stages, DARPA believes that it has shown great promise.
- **Context Aware Visualization.** The premise in this thread is that there must be additional ways developed to visualize the information for human users other than text-based lists, tables, and long passages of unstructured text. Because TIA could serve a large range of users with various roles, the visualization concepts need to be adaptive to take into account the context of the particular part of the problem being worked as well as the level of the user in the network. For example, the policy decision-maker needs different views of the information than the intelligence analyst. However, all views are based on the same underlying data and information, and even the policy decision-maker needs to be able to drill down to underlying detail periodically so that he or she has confidence in the results an analyst has provided. Different visualizations are needed for different types of analysis and the different styles and preferences of the users. The objective here is to make the information more understandable in a shorter time and by viewing data in new ways to help reveal undetected information such as patterns of activities that may otherwise be detected only by an experienced analyst.

- **Decision Making with Corporate Memory.** DARPA’s premise is that the policy decision-maker needs a wider range of actionable options earlier in the process before some options become unavailable. To make an informed decision, the policy decision-maker should have an understanding of what has happened in the past (corporate memory or a “knowledge base”) as well as an understanding in breadth and depth of the plausible outcomes of the current situation including a risk analysis of the various actionable options. The system/network (humans with the assistance of machines) should provide the policy decision-maker with information on these very complex issues, which is delivered in a manner that is quickly understandable.

Program experimentation and evaluation is taking place in two distinct channels: one is operational at the network level, and the other is R&D at the component level:

- **Operational testing of TIA network:** The premise for these activities is that the Government already possesses the data to counter terrorist threats effectively, but needs to work together better. TIA network provides an environment—an R&D prototype—for improving the analysis process. The goal in this context is to empower the individual analyst with better tools for collaboration, prediction, modeling, and database access and query while protecting the privacy of sensitive sources and methods. If successful, analysts will spend less time looking for critical information and have more time to understand the meaning of key information and have better tools for developing options for dealing with it and communicating the findings to decision-makers. These experiments and evaluations are being conducted by operational users of the participating agencies and commands while working on real-world problems using data that is legally available to them under existing laws, regulations, and policies applicable to that agency or command.
- **R&D testing of potential TIA components:** The hypothesis being tested in these activities is that it would be highly beneficial to supplement access to existing government data with access to transaction data not already in government databases. This research uses synthetic data and/or data that is legally available to the foreign Intelligence Community. The synthetic data is generated by creating imaginary people and having them do everyday things such as calling other imaginary people, traveling places, and buying things. The population of DARPA’s imaginary world is about 10 million people. Billions of transactions are being generated for this research database. This simulates normal world activity. Transactions that represent suspicious but innocent patterns of activity are inserted into this database. Finally, research teams simulating terrorist organizations are planning simulated attacks against the United States and identifying what transactions they would need to make to carry out a simulated attack. These transactions are added to the research database. University and commercial contractors are using this research database to conduct experiments to determine whether they can separate out the simulated terrorist activity from the simulated normal world activity. If this research is successful, an evaluation of the legal and policy implications would be conducted in the context of the potential benefits in preventing future terrorist attacks.

It is important to understand that the TIA vision does not include the creation of dossiers on U.S. citizens nor does it include a grand database of all U.S. transactions. DARPA wants to emphasize that no such thing is being contemplated or being implemented. Early presentations by DARPA concerning the TIA Program may have been misinterpreted. DARPA surmises the confusion arose from not distinguishing among visions on the enormity of the problem, research directions, and operational experiments. Also, some ignored the concept of filtering functions, which aim to limit the searches and pattern-based queries to only those associated with terrorist planning and preparation activities.

TECHNICAL APPROACH: TIA consists of three fundamental concepts:

- A network with an imbedded security layer to ensure that security and privacy policies are enforced that enables the sharing of data when consistent with policy. This network enables sharing of data among the intelligence, counterintelligence, national security policy, and military operations domains. For the network experiments, this is being done over a virtual private network that operates over one of the DoD physical networks for classified data.
- A secure collaborative analysis environment that allows for the ad hoc creation of intelligence and counterintelligence analyst groups that can work on common problems, postulate competing hypotheses about terrorist activities, and expose explicit evidence that supports these postulated activities in structured arguments. These competing hypotheses are passed to similar groups of operations and policy analysts who can develop estimated plausible outcomes of the current situation, actionable options, and risk analyses to be sent forward to the decision-maker.
- Numerous and reconfigurable software tools for use by the analysts in the network to quickly identify relevant data from multilingual foreign intelligence sources and to process this data for discovery of individuals, their links (relationships) to others, and their associations with groups that are related to terrorist activities. A systems view of these various stages of analysis is shown in Figure A-1.

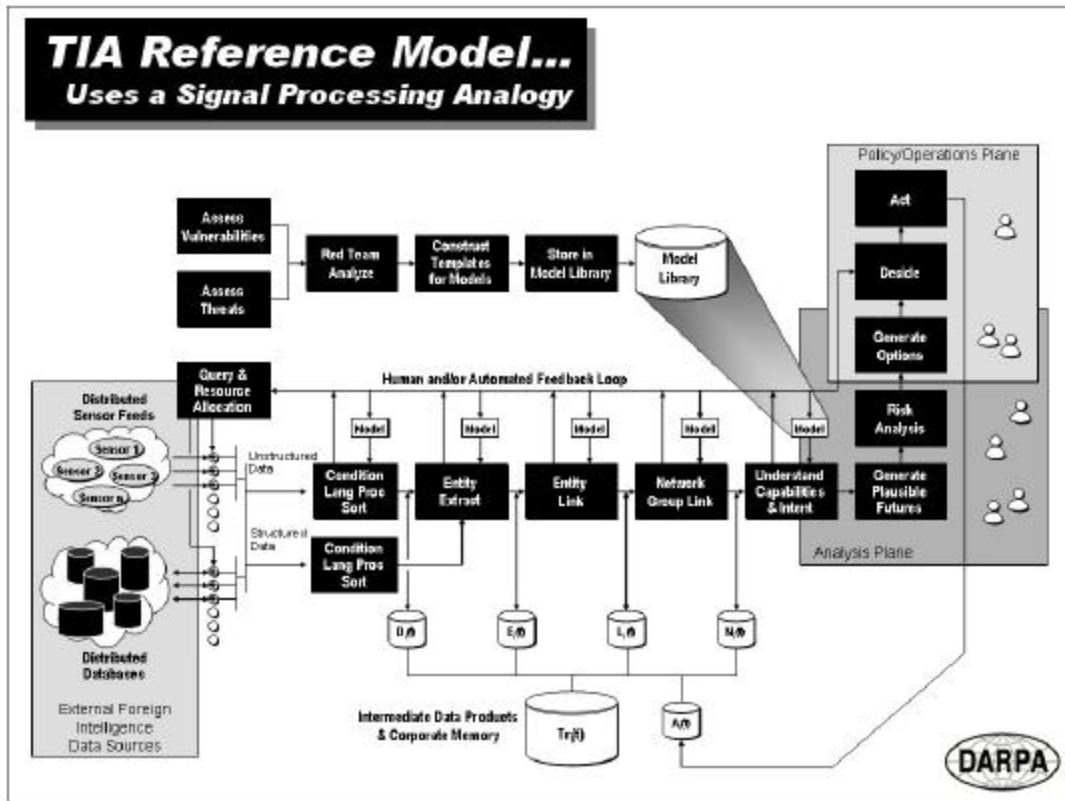


Figure A-1 - TIA Reference Model

The problem of discovering the plans and intentions of potential terrorist activities is complex. TIA is based on the premise that several key information exploitation steps must be addressed:

- Based on known vulnerabilities of the United States at home and abroad to terrorist attacks and the known and estimated capabilities of the terrorist organizations, scenarios would be developed. The planning and preparation activities to carry out these attacks would be estimated taking into account the adaptations the terrorist would most probably make to counter our defenses. Those activities that may be observable as various kinds of data in the government databases available to the intelligence communities would be converted into subject- and pattern-based queries. This information would be pulled together into a model of a terrorist attack and made available to analysts.
- Using these models and other intelligence information as starting points, analysts would initiate automated searches of their databases. These models would be refined as additional information is obtained.

- Because of the enormous amount of data already available to the government from unclassified and classified sources, automated means of processing this data and converting it to relevant information would be a monumental task beyond the capabilities of the analysts without significant new applications of information technologies.
- Individuals suspected of involvement in terrorist activities would be identified through their physical presence or the transactions they make.
- Associations among such individuals and other key entities (e.g., other people, activities, events, transactions, and places) would be made.
- These associations would be linked with the associations of other individuals.
- Other types of intelligence would be melded into the developing picture of what is happening and false leads would be identified.
- The analyst would develop hypotheses about what these associates may be planning.
- The behavior and activities of these associates may be introduced into models that are based on known patterns of behavior and activity that have been shown to be accurate or estimated to be predictors of terrorist attack.
- Based on these competing hypotheses, a range of plausible outcomes would be estimated and actionable options would be developed that address the maximum range of these plausible futures.
- A risk analysis would be done before the situation is presented to the decision-maker as early as possible so the decision-maker would have the maximum number of options to aid in deciding on a course of action or nonaction.
- All the steps of this process would be recorded faithfully in a corporate memory (knowledge base) that would be helpful in the future in similar situations.

The TIA reference model shown in Figure A-1 shows how the software components being developed in other IAO and government programs and from commercial sources fit together to provide the analysts with the capability to carry out the steps described above.

TIA is developing a system and network infrastructure, largely based on commercial standards such as those of the Internet and web-based services, that uses existing DoD communications networks and databases available to the intelligence, counterintelligence, operations, and policy communities under existing laws, regulations, and policies. This infrastructure will provide for the necessary secure collaborative environment that allows analysts on the edges of the organizations to quickly form ad hoc groups in virtual spaces (somewhat like chat rooms on the Internet) across organizational boundaries and, at the same time, use the center-based systems of their parent organizations. The secure collaborative environment will be a big step forward in

punching holes in the existing organizational “stove pipes.” The infrastructure will also provide the means for quickly plugging in new components for processing information as they become available from whatever source.

RELATIONSHIP TO OTHER IAO PROGRAMS: TIA provides an information systems architecture for a counterterrorism network and the infrastructure to support that network as described above. It also provides an experimental process in an operational environment for evaluating network performance and the efficacy of network components derived from other IAO programs, as well as other candidate commercial and government technologies—in this sense, TIA is a program of programs.

Potential contributions from other IAO programs as they relate to the various stages of the TIA reference model include:

- New Sources of Data: Biometrics-based human identification to recognize individuals and activities, with or without disguise (HumanID, NGFR, and ARM)
- Conditioning, Language Processing and Sorting: Machine translation of foreign languages, speech-to-text transcription, text summarization, semi-structuring text, and sorting by categories (TIDES, EARS, GALE)
- Evidence Extraction, Entity and Group Linking: (EELD, BioALIRT – early warning and algorithms only, MInDet, SSNA)
- Understanding Capability and Intent: (Genoa II and WAE)
- Generating Plausible Futures: Predictive modeling to estimate and predict plausible outcomes (Genoa II, RAW, FutureMAP)
- Risk Analysis and Generating Options: (Genoa II, WAE and RAW)
- Data Access and Large Semi-structured Databases: Ability to rigorously control access and query existing distributed heterogeneous government databases as if they were centralized and to maintain current and prior knowledge so possible future states can be evaluated from the perspective of an historical time continuum (Genisys)
- Privacy Protection: Ability to protect the privacy of sensitive intelligence sources and methods as well as the privacy of innocent persons (Genisys Privacy Protection)

TRANSITION/DEPLOYMENT PLANS: There are a number of organizations working with TIA in testing and evaluating the technologies under development. Organizations participating or planning to participate in TIA’s spiral development and experiments include:

- U.S. Army Intelligence and Security Command (INSCOM)
- National Security Agency (NSA)
- Defense Intelligence Agency (DIA JITF-CT)
- Central Intelligence Agency (CIA)

- DoD Counterintelligence Field Activity (CIFA)
- U.S. Strategic Command (STRATCOM)
- Special Operations Command (SOCOM)
- Joint Forces Command (JFCOM)
- Joint Warfare Analysis Center (JWAC)

All are potential transition partners if experiments are successful.

A full TIA prototype will not be ready until FY 2007; however, incremental transition of some components will take place as the components prove valuable to the user organization.

Genisys

OVERVIEW: The goal of Genisys is to make databases easy to use to increase the probability that the Government will have the information it needs. In DARPA's view, current database technology is too complex and too inflexible to represent everything we know. As a result, DARPA believes that we don't have enough automated systems and that, because we cannot keep track of so many details manually, we lose control of critical information.

The Genisys Program is conducting R&D to make it easier to integrate the information in existing databases used by different agencies involved in counterterrorism so they can share what they know and correlate events faster. Current technology for integrating databases is slow, tedious, and error prone. To integrate databases or even to use the information they contain, analysts first have to know that databases exist and they must have access permissions, which are normally supplied manually (i.e., slowly) by a system administrator. Symbols used inside the database must be interpreted by analysts who may be unfamiliar with the domain and are probably uncertain about the precise meaning of terms. To execute complex queries, analysts must also know a great deal about the database design and how to join different sets of information to get the right answers. As a result, it is very difficult to access information from disparate sources.

Another related goal that DARPA has for the future is developing and evaluating technology to enable very large databases as a foundation of knowledge about terrorists for preventing future attacks. DARPA believes that to predict, track, and thwart attacks, the United States needs databases containing information about all potential terrorists and possible supporters; terrorist materials; training, preparation, and rehearsal activities; potential targets; specific plans; and the status of our defenses. In DARPA's view, current database technology is not adequate to meet these needs, and the Genisys Program seeks to fix that problem.

DARPA believes that current commercial database technology is inadequate to meet the large-scale massive data needs for countering terrorism. Today's database technology was defined in the 1970s, but processors, disks, and networks are now thousands of times more capable. Genisys seeks to reinvent database technology to meet today's counterterrorism needs and capabilities. Genisys would also stress-test research ideas by developing a series of increasingly

powerful leave-behind prototypes so the Intelligence Community can get value immediately and provide feedback to focus research. When developed, these technologies and components would be evaluated for applicability to TIA.

TECHNICAL APPROACH: The Genisys Program’s technical approach includes integrating databases and other information sources using mediation and representing uncertain information explicitly using probability bounds. The prime contractor for Genisys, AlphaTech, and its subcontractor, Oracle, aims to create technology that enables many physically disparate heterogeneous databases to be queried as if it were one logical “virtually” centralized database. The technology, mediation, refers to the use of intelligent software agents that would relieve analysts from having to know all the details about how to access different databases, the precise definitions of terms, the internal structure of the database and how to join information from different sources, and how to optimize queries for performance. Instead, this information would be encoded and managed by software agents. As a result, analysts would be able to access information much faster and with higher confidence in their results. They would be able to use all the databases to which they have access as a federation—a new “megadatabase” would not be created. Information from other sources such as the web or semi-automated collection systems would be somewhat easier to convert into structured data and that would help TIA increase its information coverage. Finally, the developers seek to create a probabilistic database engine for representing and dealing with uncertainty. The development efforts will extend over 5 years, but mature increments will be delivered annually.

Genisys and related TIA efforts will make significant use of synthetic data to support development and testing. The premise underlying TIA’s focus is that information systems and databases have unique potential in identifying terrorist planning and preparation activities through the transactions they make. However, Americans are rightfully concerned that data collection, analysis, and mining activities by intelligence analysts threaten their privacy. In order to evaluate the usefulness of TIA technologies, a smaller research effort related to Genisys is creating a realistic world of synthetic transaction data using intelligent software agents that simulate the behavior of normal people, unusual-but-benign people, and terrorists. Grounded in the physical world but populated by imaginary people, the resulting transaction database will allow researchers to explore new technologies for detecting, identifying, tracking, and elucidating synthetic terrorist attacks using unclassified data that is not contaminated with transaction data about real U.S. persons.

DARPA is using a red team of terrorism experts who are creating synthetic terrorist attack scenarios and will produce transaction data reflective of these attacks. This manually generated data will be rational, innovative, and devious. As the simulation and the pattern-matching technology improves, the experts on the red team will be able to create richer and more sophisticated attack scenarios, and the query and pattern detection software researchers will be constantly challenged.

DARPA’s goal for this activity is to find out what is possible. If the concepts and algorithms in IAO programs cannot extract terrorist signatures from a world of noise, even for simulated data, there is no reason to proceed. However, if the technology works in a realistic simulation, its advantages for protecting the Nation against terrorism can be weighed against its potential for reducing personal privacy.

RELATIONSHIP TO TIA: The prototype capabilities from each phase of Genisys will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each Genisys component.

TRANSITION/DEPLOYMENT PLANS: Genisys software components will be evaluated in a series of TIA experiments beginning in FY 2004. Based on the results of these experiments, successful Genisys technology will be transitioned as applicable.

Genisys Privacy Protection

OVERVIEW: The Genisys Privacy Protection Program will research and develop new technologies to ensure personal privacy and protect sensitive intelligence sources and methods in the context of increasing use of data analysis for detecting, identifying, and tracking terrorist threats. Information systems and databases have the potential to identify terrorist signatures through the transactions they make. Americans are rightfully concerned that data collection and analysis activities by the Intelligence Community threaten their privacy. To address this concern, the Genisys Privacy Protection Program will conduct R&D on technologies that enable greater access to data for security reasons while protecting privacy by providing critical data to analysts while not allowing access to unauthorized information, focusing on anonymized transaction data and exposing identity only if evidence warrants and appropriate authorization is obtained for further investigation, and ensuring that any misuse of data can be detected and addressed.

If successful, Genisys Privacy Protection will develop privacy algorithms that prevent unauthorized access to sensitive identity data using statistical and logical inference control. This privacy protection technology would be used to develop roles-based rules for distinguishing between authorized and unauthorized uses of data and will automate access control. The program will also seek to improve the performance of algorithms for identity protection by limiting inference from aggregate sources. DARPA's research activities under the Genisys Privacy Protection Program include the development of mechanisms and a trusted guard for access control and immutable audit that would be available to an appropriate oversight authority. This appliance would enable methods to automate audit, identify potential privacy violations, and uncover underlying goals and information content from obscure and distributed query sets.

Access to Government databases today is granted ad hoc by system administrators. Thus, access is nonstandard, slow, and often not granted unless direct interaction is mandated. Terrorists have already exploited the inability to share information and act collaboratively on problems. Role-based access control using standardized business rules would automate access appropriately, in a controlled and well-understood manner. To track information that leaves the database, DARPA has identified an innovation called "self-reporting data"—data that, when accessed, reports its location and the person accessing it to an automated information tracking system. This technology may have utility not only for personal privacy, but also for the intelligence insider threat.

Genisys Privacy Protection would permit analysis of critical data while protecting sensitive information such as personal identity or sources and methods. Filters and software agents would be used to eliminate any data that is not potentially useful for combating terrorism. All transactions would be anonymized prior to analysis; that is, information that implies personal identity would be removed and, in general, less sensitive data would be analyzed until patterns match and more sensitive data is justified to test hypotheses.

To test these ideas, DARPA is examining the feasibility of a privacy appliance—hardware and software that sits on top of a database, which is controlled by some appropriate oversight authority, and has mechanisms to enforce access rules and accounting policy. The idea is that this device, cryptographically protected to prevent tampering, would ensure that no one could abuse private information without an immutable digital record of their misdeeds. The details of the operation of the appliance would be available to the public. Methods such as encryption and distribution would protect both ongoing investigations and the possibility of covering up abuse.

TECHNICAL APPROACH: The Genisys Privacy Protection technical approach includes implementing component technology, controlling inference, and automating audit to increase the odds of catching any abuse. DARPA's contractor for Genisys Privacy Protection is Palo Alto Research Center (PARC, formerly part of Xerox Corporation). PARC will invent new technology for addressing the problem of combining information from several sources, none of which by themselves provide sensitive information, but that in the aggregate can result in inferences that expose more private data than was originally intended. This is a difficult technical problem to overcome because, once information is known, it can be combined with other information in an infinite number of ways and information can come from many different sources. PARC will address this problem by encoding information that is useful for aggregate analysis and continually tracking and metering information from all sources. Algorithms would perform some analysis automatically and would shut off information when human analysts exceed some "knowledge threshold" unless additional analysis is warranted and approved by the appropriate authority. In this way, analysts would be able to access information they need, but would require additional justification for information beyond a set threshold. In addition, PARC will create sophisticated algorithms to identify any unusual use of data that may indicate abuse. They would automate audits to increase the likelihood that those who might misuse their access to information will be caught.

RELATIONSHIP TO TIA: The prototype tools from each phase of Genisys Privacy Protection will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each Genisys Privacy Protection tool.

TRANSITION/DEPLOYMENT PLANS: Genisys Privacy Protection components will be evaluated in a series of TIA experiments beginning in FY 2004. Based on the results of these experiments, successful technology will be transitioned in the form of permanent components in a TIA prototype.

Evidence Extraction and Link Discovery (EELD)

OVERVIEW: Preliminary EELD activities were initiated in 1999, predating the TIA Program as well as the attack of September 11, 2001. The full program began in FY 2001. The objective of EELD is to develop technology for “connecting the dots”—starting from suspicious activities noted in intelligence reports. The EELD automated toolset, once developed, will assist intelligence analysts by automatically drawing to their attention the key relationships among subjects of lawful investigations drawn from the materials currently gathered and reported about non-U.S. persons, and it will do so in a both timely and comprehensive manner. For example, it has been widely reported in the press that the significance of a key planning meeting of Al-Qaeda in Malaysia prior to September 11, 2001, was not recognized in time for the CIA to place the participants on immigration watch lists until August 2001, which was too late to prevent the attacks because by then they had already entered the United States. EELD techniques will also be useful in reducing false alarms because they would enable the explanation of certain patterns of activity as legitimate and, therefore, unworthy of retention or investigation, separating these instances from those with no legitimate explanation or those whose participants are connected to known or suspected terrorists.

DARPA believes that EELD is needed because commercial data-mining techniques are focused at finding broadly occurring patterns in large databases, in contrast to intelligence analysis that consists largely of following a narrow trail and building connections from initial reports of suspicious activity. Commercial data-mining techniques are typically applied against large transaction databases, while intelligence needs to focus on a much smaller number of people, places, and things engaging in a far wider variety of activities. Commercial techniques attempt to sort all transactions and the people who make them into classes based on transaction characteristics; intelligence needs to combine evidence about multiple activities from a small group of related people. Patterns observed in commercial databases must be widespread to be of interest to companies; patterns that indicate activity of interest to the Intelligence Community are extremely rare. Commercial data mining combs many large transaction databases to discover predominant patterns; EELD technology combines information extracted from intelligence reports to detect rare but significant connections. The goal of the EELD research program is to extend data-mining technology and develop new tools capable of solving intelligence problems; it is not performing data mining as the term is currently understood in the commercial sector.

TECHNICAL APPROACH: EELD assumes that the initial information for analysis is currently available—but potentially unrecognized—from traditional intelligence sources, although not in a form that is easily analyzed. Information contained in these sources is collected based on initial indications of suspicion in conformance with laws, policies, and regulations governing the operation of these communities. This information is reported in regular textual documents and stored in existing information systems used by these communities. The technology requirements for EELD fall into the following categories:

- Evidence Extraction (EE): Because key intelligence information is typically reported in regular textual reports, it is necessary to extract specific information from these reports. Current technology allows for the automated accurate recognition and extraction of information about people, places, and things, but not about their connections and interactions, which is the key to successful intelligence analysis.

Therefore, EELD has as its first technology goal the ability to extract information from textual documents about relationships among people, places, and things. Current technology for extracting facts from text may be thought of as focused at nouns and adjectives; EELD's extraction technology will add the capability to extract information expressed by verbs and adverbs.

- **Link Discovery (LD):** Information extracted from intelligence reports about suspicious people, places, and things and their connections can be placed in a database where it can be connected to other related information. These additional connections may indicate increased significance of the information. The significance of the connected information can be recognized by noticing its connection to previously known suspicious people, places, or things or its correspondence to suspicious patterns of activity. Once an initial indication of suspicion is present, a search process may be initiated of other databases available to the Intelligence Community to fill in more blanks and aid in the human analyst evaluation of the emerging information. Human experience and judgment combined with this additional information allows an analyst to determine if the apparently suspicious information is explainable as an example of unusual but legitimate activity or if further investigation is warranted. If further investigation is warranted, it would be undertaken according to the policies and procedures governing the particular agency. LD techniques and tools support this process of making connections, evaluating significance, searching for additional related information, recognizing potential patterns of interest, and eliminating explainable patterns in this mass of connected data. LD is the core of EELD; it is the technology for "connecting the dots."
- **Pattern Learning (PL):** Initial patterns that indicate potentially suspicious activity come from experienced intelligence analysts. However, there may be suspicious activities that have not previously occurred, but are still appropriate and worthwhile to investigate. Also, patterns of interest may change over time as potential adversaries adapt their behavior or as new types of legitimate activities occur. PL is aimed at developing technology to increase the accuracy of patterns in discriminating between suspicious and legitimate activity, to suggest previously unknown but potentially significant patterns, and to adapt known patterns as adversarial and legitimate behaviors evolve.

RELATIONSHIP TO TIA: The prototype tools from each area of EELD will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each EELD tool.

TRANSITION/DEPLOYMENT PLANS: EELD is developing technology—conducting research by developing algorithms, implementing these algorithms in software, evaluating these algorithms individually and in combination on carefully engineered test problems, and integrating the useful and effective algorithms into software tools that can be provided to the Intelligence Community or to system developers. EELD technology developments use and are regularly evaluated against both open source and simulated data with characteristics and properties carefully engineered to match scenarios of interest, but with fictitious individuals and with a controlled variation of parameters to enable effective and valid experimentation.

Particular software tools developed in the EELD Program have been incorporated into TIA experiments with classified intelligence data where their value to the Intelligence Community is being established. EELD funds system concept and performance assessment activities to ensure that specific technologies are applicable to the types of data available and to the analysis processes used by the Intelligence Community; to enable joint experimentation with combined technologies; to construct test problems (consisting of data sets containing examples of both suspicious and legitimate activities, specifications of patterns of suspicious and legitimate activities, and answer keys); to conduct experiments and evaluations; and to enable transitions to user organizations.

Scalable Social Network Analysis (SSNA)

OVERVIEW: The SSNA Program is developing techniques based on social network analysis for modeling the key characteristics of terrorist groups and discriminating these groups from other types of societal groups. Social network analysis (SNA) techniques have proven effective in distinguishing key roles played by individuals in organizations and different types of organizations from each other. For example, most people interact in several different communities; within each community, people who interact with a given individual are also likely to interact with each other. Very preliminary analytical results based on an analysis of the Al-Qaeda network of September 11 hijackers showed how several social network analysis metrics changed significantly in the time immediately prior to September 11; this change could have indicated that an attack was imminent. Current SNA algorithms are effective at analyzing small numbers of people whose relationship types are unspecified; SSNA would extend these techniques to allow for the analysis of much larger numbers of people who have multiple interaction types (e.g., communication and financial). The program will develop algorithms and data structures for analyzing and visualizing the social networks linkages, implement algorithms and data structure into software modules that provide SNA functionality, and demonstrate and evaluate these models in appropriate Intelligence Community systems and environments.

TECHNICAL APPROACH: SSNA will develop scalable algorithms and the data structures essential to support the analysis of social networks comprising large numbers of individuals who may be linked by a multitude of interactions. The program will explore techniques in graph theory, SNA, and mathematics to identify networks of multiple relationships among individuals and/or organizations in open source materials. It will be necessary to create the ability to analyze data structures to characterize the social network as being an abnormal or a normal SNA network. Ultimately, the program will strive for the capability to illustrate SNA network activities evolving from a dormant to an active stage over time.

RELATIONSHIP TO TIA: The prototype tools from each phase of SSNA will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each SSNA tool.

TRANSITION/DEPLOYMENT PLANS: Organizations with a strong potential for using SSNA technology include the Defense Intelligence Agency (DIA), CIA, National Security Agency, and other military commands with requirements for intelligence analysis or for ensuring force protection and national security.

MisInformation Detection (MInDet)

OVERVIEW: The objective of the MInDet Program is to reduce DoD vulnerability to open source information operations by developing the ability to detect intentional misinformation and to detect inconsistencies in open source data with regard to known facts and adversaries' goals. Open source information may exist in news reports, web sites, financial reports, maritime registrations, etc. By its very nature, it is public information. At present, the Intelligence Community does not take full advantage of open sources, for a number of reasons. One reason is because of the sheer volume of open source information. Another key reason is the lack of reliability of open sources. The motivating idea for MInDet is that automated determination of reliability of open sources will allow U.S. Intelligence to fully exploit these additional sources. Techniques will be developed for detecting misleading information in single documents, such as visa applications or maritime registrations as well as in a series of reports, e.g., news reports from different sources in a foreign country.

Five Small Business Innovation Research (SBIR) efforts were conducted in FY 2002 to explore preliminary ideas regarding feasibility of different technical approaches. Three SBIR efforts are planned to continue during FY 2003 to further develop the more promising approaches. Intelligence Community experts are participating in the development of the detailed program plan and approach.

TECHNICAL APPROACH: MInDet will develop domain-specific indicators of potential intentional misinformation in open source material using red team wargaming techniques and expert knowledge. The program will explore combinations of techniques from linguistic genre analysis, learning with background knowledge, business process modeling, and adversarial plan recognition for detection of intentional misinformation in open sources. MInDet seeks promising algorithms using a number of approaches (such as combination of linguistic processing, knowledge-based reasoning, and Bayesian inferencing; decision-tree approach to detect red-flag conditions associated with creative financial reports; deductive anomaly detection; Bayesian technique for evidence fusion; and categorization and concepts extraction) to detect misinformation. The benefits of this technical approach and of MInDet tools will be assessed by demonstrating the ability to detect misinformation in a number of domains such as detecting inconsistencies in news releases between internal and external consumption, classification performance of detection effectiveness and computational resources from known fraudulent/suspicious company websites, and detecting red-flag conditions in Security and Exchange Commission (SEC) filings. In FY 2003, the MInDet Program is continuing with three Phase II SBIR efforts to implement conceptual prototypes for the concepts that were validated during Phase I. These prototypes will detect misinformation in semistructured data (such as web pages) in large volumes of semistructured documents and data streams; 2) use domain-independent deception heuristics and information extraction techniques against a diversity of

evidence sources; and 3) use linguistic techniques to identify features that serve as indicators of misinformation in multilingual sources. Also during FY 2003, we will construct challenge problems, consisting of large sets of open source information, some of which are examples of misinformation. A competitive solicitation will be conducted to select the research approaches and performers for the full program.

RELATIONSHIP TO TIA: The prototype tools from each phase of MInDet will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each MInDet tool.

TRANSITION/DEPLOYMENT PLANS: Organizations with a strong potential for using MInDet technology include the DIA, CIA, the National Security Agency, and other military commands with requirements for terrorist threat detection, national security, intelligence analysis, and information operations. There are also very strong potential applications for use in law enforcement and regulatory applications. As spelled out in the Report, careful consideration would have to be given to a number of factors before deployment in such latter contexts.

Human Identification at a Distance (HumanID) Program

OVERVIEW: The HumanID Program predates both the Information IAO and the TIA Program. The goal of HumanID is to develop automated biometric recognition technologies that will enhance force protection and provide early warning against terrorist and other human-based threats. Obtaining this information can decrease the effects of or prevent such attacks and provide more secure protection of critical military and operational facilities and installations.

HumanID plans to accomplish this goal by developing biometric technologies that are able to detect, recognize, and identify humans using a combination of biometric modes at distances up to 500 feet. Biometric technologies being developed include face, gait, and iris recognition. Digital, video, infrared and hyperspectral imaging technologies are also being investigated.

Once these individual technologies have been developed and assessed, HumanID will develop methods to fuse the most promising technologies with the intent of increasing the performance, reliability, and range of applications. To conduct HumanID research, biometric signatures will be acquired from various sensors including video, infrared, and multispectral sensors under varying conditions or scenarios. Experiments and evaluations will be conducted using these signatures to determine the most promising approaches.

TECHNICAL APPROACH: Today, most face recognition systems work best on frontal images taken at close range (under 10 feet), using cooperative subjects under indoor lighting conditions. To increase the range at which the face can be recognized, the HumanID Program created an active vision face recognition system. The system detects people and faces between 20 and 150 feet and then zooms in to recognize the detected face. The HumanID Program has worked on increasing face recognition performance. In 2 years, the program has reduced the

error rate on recognition from frontal indoor images by 50 percent. The development of three-dimensional morphable models has greatly increased the capability to recognize nonfrontal faces. The HumanID Program is also investigating experimentally benchmarking human performance under different viewing and noise conditions. This will allow the comparison of human and machine capabilities. This knowledge can be used to design better algorithms and human-computer interfaces. The program will also attempt to create a large sample of spontaneous behavior that will provide the basis for testing and perfecting robust techniques of face recognition and for providing information on behaviors under stressful deception situations and human emotion expression. Finally, HumanID will develop and evaluate models and algorithms for human identification of freely behaving individuals with natural interaction and spontaneous expressions.

The performers in the body dynamics area of investigation are conducting research to determine if the way a person moves and walks (gait) is a unique and identifiable biometric that can be used in detecting and recognizing a human. The majority of this effort is the gait recognition challenge problem. Gait recognition approaches from six different universities were compared against a common baseline approach, which allows for the assessment of the best approaches and common strengths and weakness of all gait algorithms. The conclusions to date are that gait fused with face has the potential to improve face recognition performance; gait can improve the reliability of tracking algorithms; and gait is a fundamental component for future research in human activity inference. Other areas under investigation are techniques for fusing gait and face for identification, improved methods for locating humans in video, and improved methods for human activity inference.

Performers in the sensors area of research are developing advanced biometric sensors and signal and image processing techniques to determine if unique biometric signatures exist and whether they can be used to detect and recognize individual humans. Technology being researched includes sensors to improve human identification in bad weather and from thermal infrared imagery, a sensor that can recognize the iris of a cooperative individual from up to 10 meters away, radar technology capable of recognizing an individual in a crowd of less than 100 people, sensors and techniques with the ability to identify people in a 360-degree view, techniques for using an individual's physiological features to identify them, and algorithms for identifying people from hyperspectral images acquired under uncontrolled conditions.

To assess the performance of the biometric technologies being developed, large-scale tests and evaluations are necessary. The HumanID Program was a sponsor of the Face Recognition Vendor Test (FRVT) 2000 and FRVT 2002, large-scale evaluations of core face recognition technology sponsored by numerous Government agencies. The evaluations provide an assessment of the state of the art and identify future research directions. For prototype systems, in-situ field demonstrations/evaluations are performed. To advance understanding of how systems work and to more accurately measure performance, advanced statistical techniques for measuring performance are being developed. Because a large amount of data is required to develop new biometrics technologies and assess their performance, some of the HumanID performers in this research area are acquiring biometric data sets under carefully controlled conditions (described below). The complete corpus of biometric signatures collected under the HumanID Program is called the *hbase*. The *hbase* allows for systemic experimentation, testing,

and evaluation of biometric techniques and algorithms developed under the HumanID Program. All biometric collection activities are vetted through each participating institution's internal review board (IRB) or process to ensure compliance with human subjects regulations. Participation in the biometric collection is completely voluntary, and biometric signatures are stored anonymously. Each participant is given a random identifier with no record linking the identifier to the person.

RELATIONSHIP TO TIA: At this time, the HumanID Program does not have specific technology to transition to TIA. Improvements in face recognition algorithms are incorporated in the next generation of commercial-off-the-shelf face recognition systems. The program is working with the Office of Naval Research (ONR) to jointly demonstrate and evaluate HumanID technology for protection of port facilities and naval vessels in port. The projected demonstration would include face recognition from visible imagery, face recognition from dual mode infrared and visible imagery, and gait recognition. The naval facility in Bahrain is of particular interest. Some HumanID technologies need further development and refinement prior to transition to the Military Services or DoD. If future HumanID research proves successful and transitions to operational use, it may provide input to TIA via distributed sensor feeds in a manner similar to existing feeds within the foreign Intelligence Community.

TRANSITION/DEPLOYMENT PLANS: There are a number of organizations working with HumanID in testing and evaluating the technologies under development. These organizations include:

- National Institute of Standards and Technology (NIST) – biometric evaluation processes and protocols
- National Institute of Justice (NIJ) – Co-sponsor, FRVT 2002. Co-funded NIST to develop biometric evaluation standards.
- DoD Counterdrug Technology Development Program Office – Co-Sponsor, FRVT 2002
- Transportation Security Administration (TSA) – Co-sponsor, FRVT 2002
- Federal Bureau of Investigation (FBI) – Co-sponsor, FRVT 2002
- CIA – Support to foreign intelligence operations
- U.S. Army (INSCOM and Natick Laboratories) – Initial HumanID demonstration; development of force protection and physical security concept of operations (CONOPS)
- U.S. Air Force (USAF Force Protection Battle Laboratory) – Second HumanID demonstration (visual and infrared recognition)
- U.S. Special Operations Command (USSOCOM) – CONOPS development

Activity, Recognition, and Monitoring (ARM)

OVERVIEW: The goal of the ARM Program is to develop an automated capability to reliably capture, identify, and classify human activities. Currently, these types of activities are identified and analyzed by humans studying real-time and recorded video sequences. ARM technology will dramatically improve the speed and ability to discover and identify anomalous or suspicious activities in DoD facilities in the United States or abroad.

The ARM Program plans to develop technologies to analyze, interpret, model, and understand human movements; individual behavior in a scene; and crowd behavior. ARM will develop human activity and scenario-specific models that will enable operatives to differentiate between normal and suspicious activities in a given area or situation. The capability to automatically identify and classify anomalous or suspicious activities will greatly enhance homeland defense initiatives by providing increased warning for asymmetric attacks, increase the reconnaissance and surveillance capabilities for Intelligence and Special Operations Forces, and provide more secure protection of critical DoD military and civilian facilities and installations.

Situations where ARM technology will significantly improve current surveillance capabilities include searching for unusual patterns of activity; and discovering unattended packages and identifying individuals who are casing, loitering, or observing critical facilities. In particular, this includes detecting hostile operatives collecting data on deployed forces, critical infrastructure components, or DoD facilities at home or abroad.

TECHNICAL APPROACH: The ARM Program will develop intelligent activity and monitoring algorithms that are resident in networked sensors; develop a scalable, extensible prototype system of networked sensors; demonstrate and evaluate the prototype system in a series of increasingly challenging scenarios; create a database capable of searching observed activities for retrospective analysis; and develop human-computer interfaces that are tailored to user demands.

RELATIONSHIP TO TIA: If ARM research proves successful and transitions to operational use, it may provide input to TIA via distributed sensor feeds in a manner similar to existing feeds within the DoD Intelligence Community.

TRANSITION/DEPLOYMENT PLANS: ARM will conduct close coordination and periodic technology assessments with military force protection elements (U.S. Army INSCOM; U.S. Air Force, Force Protection Battle Lab; and Natick Labs). Successful technologies developed under the ARM Program will transition to the Military Services as product improvements to previously fielded face recognition systems.

Next-Generation Face Recognition (NGFR)

OVERVIEW: The objective of the NGFR Program is to initiate development of a new generation of facial-based biometrics that can overcome face recognition operational challenges/scenarios and are robust to time differences between facial imagery, pose, and illumination.

DARPA believes that the importance of facial biometrics has become clear in the aftermath of the events of September 11, 2001. The two most mature facial biometrics are mug shot-style facial imagery and iris scans. The performance of face recognition from mug shot-style imagery is well understood, and areas for improvements in performance have been documented and are underway. Promising new techniques with the potential to overcome current limitations of mug shot-style face recognition are beginning to emerge. These techniques include the use of high resolution imagery, the employment of 3-D imagery and processing technologies, expression analysis, and analysis of the temporal information inherent in video imagery. Advanced facial biometrics can also provide clues to indicate if a person is being deceptive. Deception detection requires automatic identification and classification of complex facial expressions, which thwart state-of-the-art face recognition technologies as well as automatic detection of deceptive expressions, behaviors, or characteristics, which may indicate hostile intent on the part of known or unknown rogues. Expression analysis and video analysis are methods for detecting deception. A small research project in this area is establishing the foundation for developing methods and algorithms for detecting deception in visual signals.

TECHNICAL APPROACH: The NGFR Program is conducting research in four technology areas: facial feature detection, tracking, and classification to include rapid facial motion, head motion, talking, and other facial expression activity; facial identification using 3-D morphable models to counter the effects of occlusion from head motion, pose, lighting, glasses and facial hair; automatic detection of biometric indications of deceptive behavior; and characterization of iris recognition.

RELATIONSHIP TO TIA: If NGFR research proves successful and transitions to operational use, it may provide input to TIA via distributed sensor feeds in a manner similar to existing feeds within the DoD Intelligence Community.

TRANSITION/DEPLOYMENT PLANS: NGFR will conduct close coordination and periodic technology assessments in conjunction with TIA development efforts at INSCOM and at military force protection elements via the U.S. Air Force, Force Protection Battle Lab and Natick Labs. Successful technologies developed under the NGFR Program will transition to the Military Services as product improvements to previously fielded face recognition systems.

Appendix B – Other IAO Programs

Programs that may provide technology as possible components of a TIA prototype but are considered of secondary interest within the context of this report are:

- Genoa II
- Wargaming the Asymmetric Environment (WAE)
- Rapid Analytical Wargaming (RAW)
- Futures Markets Applied to Prediction (FutureMAP)
- Effective, Affordable, Reusable Speech-to-Text (EARS)
- Translingual Information Detection, Extraction and Summarization (TIDES)
- Global Autonomous Language Exploitation (GALE)
- Babylon
- Symphony
- Bio-Event Advanced Leading Indicator Recognition Technology (Bio-ALIRT)

Genoa II

OVERVIEW: As the nation faces the terrorist threat, different groups of people will need to work together as teams. Team members may be drawn from local, state, and federal governments. Members may represent law enforcement, intelligence, policy, decision-making, and operational organizations. Most—if not all—of the time, individuals will not be collocated. Team members need to work effectively within their own organizations while simultaneously supporting the team. Team members will join and leave teams as situations and resources demand. Even in such a challenging environment, teams must function with peak efficiency.

Genoa II will provide collaborative reasoning tools for TIA that will enable distributed teams of analysts and decision-makers to more effectively use the information resources available. The impact will be more rapid processing of incoming data, more complete analysis of possible hypotheses, more accurate understanding of complex situations, more accurate understanding of possible future situations, and more optimal selection of decision options.

The goal of Genoa II is to develop collaboration, automation, and cognitive aids technologies that allow humans and machines to think together about complicated and complex problems more efficiently and effectively. The project will develop technology to support collaborative work by cross-organizational teams of intelligence and policy analysts and operators as they develop models and simulations to aid in understanding the terrorist threat, generate a complete set of plausible alternative futures, and produce options to deal proactively with these threats and scenarios. The challenges such teams face include the need to work faster; overcome human cognitive limitations and biases when attempting to understand complicated, complex, and uncertain situations; deal with deliberate deception; create explanations and options that are persuasive for the decision-maker; break down the information and procedural stovepipes that existing organizations have built; harness diversity as a tool to deal with complexity and uncertainty; and automate that which can effectively be accomplished by machines so people

have more time for analysis and thinking. Emphasis will be on ease of use, adaptation to the user who is often not a scientist or engineer, and implicit encouragement to use the tools to make the users' tasks easier.

Genoa II will strive to develop innovative technology for automating some of the team processes; augmenting the human intellect via tools that assist teams thinking together, tools that do some of the thinking for people, and tools that support human-machine collaboration in the cognitive domain; and for providing a rich environment for collaboration across existing hierarchical organizations while maintaining the necessary accountability and control. DARPA envisions that the human teams using these tools will be drawn from multiple organizations spanning state, local, and federal governments. Thus, there will be the need to permit collaboration across organizational boundaries while providing control and accountability and connection back to the central systems of each participating organization. Technology will be required to support the entire life cycle of such teams. Key challenges include knowledge management/corporate memory, declarative policy generation and context-based enforcement, business rules and self-governance, and planning and monitoring team processes.

The goals for automation technology include speeding the front-end processes of gathering, filtering, and organizing information and assimilating its content without having to read all of it. On the back end of the process, technology is needed to automate or semi-automate the generation of efficient and persuasive explanations and to maintain consistency within a large, distributed multimedia knowledge base. Technology is required to make the tools and the collaborative environment itself more efficiently used by humans by making it aware of user context and preferences and smart and adaptive to optimize the user experience. There is a need for technology to aid the human intellect as teams collaborate to build models of existing threats, generate a rich set of threat scenarios, perform formal risk analysis, and develop options to counter them. These tools should provide structure to the collaborative cognitive work and externalize it so it can be examined, critiqued, used to generate narrative and multimedia explanations, and archived for reuse.

TECHNICAL APPROACH: Genoa II will address these needs by developing new information technology in three broad areas:

- Evidential Reasoning, Scenario Generation, and Explanation. This area includes the development of structured argumentation and evidential reasoning tools that will help the analyst organize available data; generate hypotheses to understand the current situation; generate possible futures that might develop from the current situation; generate and analyze possible interdiction options; and generate explanations of the analysis and reasoning process for decision-makers.
- Collaboration and Corporate Memory. This area includes the development of computing infrastructure to enable distributed teams of analysts and decision-makers to form teams, share information, and collaborate throughout the evidential reasoning, scenario generation, and explanation process. This technology needs to support collaboration at the "edge" of very different organizations while simultaneously

allowing “edge-to-center” collaboration between individual members of these groups and the “center” of their home organizations.

- Read Everything (Without Reading Everything). This area includes the development of technology to help the analyst internalize and understand all the available information relevant to understanding the current situation without having to read all of it.

These technologies will be developed and evaluated in three major phases:

- Edge-Based Collaboration for Argument Construction. During the first 18 months of the program (1st Quarter FY 2003 through 2nd Quarter FY 2004), a basic suite of evidential reasoning, collaboration, and read-everything tools will be developed and evaluated. The evidential reasoning tools will provide the basic capability for analysts to construct, reason about, and explain structured arguments. The collaboration component will provide a basic peer-to-peer collaboration capability for edge-to-edge organizational components to form and manage ad hoc teams. The read-everything tools will provide basic information retrieval capabilities.
- Center-Edge Collaboration for Evidential Reasoning and Scenario Generation. During the next 18 months (3rd Quarter FY 2003 through 4th Quarter FY 2005), an enhanced suite of tools will be developed and evaluated. The evidential reasoning component will be enhanced to include tools for hypothesis comparison, argument critique, analogical reasoning, scenario generation, stochastic option generation, and storytelling. The collaboration component will be enhanced tools to provide an initial center-edge collaboration environment, which will include context-based business rules, workflow management, SNA-based team management, consensus analysis, and knowledge-based security filters. The read-everything tools will provide alternative techniques for detecting and tracking content changes, relevant to the analyst’s situation, in the incoming data streams.
- Full Center-Edge Integration. During the last 2 years of the program (1st Quarter FY 2006 through 4th Quarter FY 2007), a full center-edge collaboration environment with a full suite of evidential reasoning, scenario generation, and explanation capabilities will be developed and evaluated.

RELATIONSHIP TO TIA: Genoa I, the predecessor to Genoa II, is already providing early tools in the structured argumentation area. The prototype tools from each phase of Genoa II will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each Genoa II tool.

TRANSITION/DEPLOYMENT PLANS: The Genoa II Program seeks a rapid, major leap in technology supporting cross-organizational teams of intelligence and policy analysts and operators working the terrorist threat. The goal is to enable teams to make much more effective use of available information. If successful, the tools will afford more rapid processing of

incoming data, more complete analysis of possible hypotheses, more accurate understanding of complex situations, more accurate understanding of possible future situations, and more optimal selection of decision options.

The Genoa II Program will utilize the talents of 11 contractor teams and will build on earlier work sponsored by DARPA and others. The program is planned as a 5-year spiral development effort in which test and evaluation of early prototype tools will guide subsequent work. In the first year, contractor teams will research and prototype information tools designed to support teams. All contractors will develop appropriate metrics against which to measure the individual prototype tools under development. As they are developed, prototype tools will be passed to the TIA test and evaluation contractor for comprehensive evaluation. The results of all testing and evaluation will guide the future development of tool functionality and integration in subsequent years. Similar prototype development and evaluation cycles will occur throughout the 5-year program. Genoa II emphasis is research and proof of concept through the development and evaluation of prototype tools.

Transition decisions for the tools developed under the Genoa II Program will be made by the TIA Program.

GENOA II - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$10,501	\$20,403	\$19,910	FY 2007

PROGRAM SCHEDULE: Genoa II began in FY 2003 and will conclude in FY 2007. The current schedule follows.

Milestone	FY/Quarter
Evaluate Phase I Evidential Reasoning Components	FY03 (4Q)
Evaluate Phase I Collaboration Components	FY03 (4Q)
Evaluate Phase I “Read Everything” Components	FY03 (4Q)
Software Drop #1 to TIA System	FY04 (1Q)
Evaluate Phase II Evidential Reasoning Components	FY04 (4Q)
Evaluate Phase II Collaboration Components	FY04 (4Q)
Evaluate Phase II “Read Everything” Components	FY04 (4Q)
Software Drop #2 to TIA System	FY05 (1Q)
Evaluate Phase III Evidential Reasoning Components	FY05 (4Q)

Milestone	FY/Quarter
Evaluate Phase III Collaboration Components	FY05 (4Q)
Evaluate Phase III “Read Everything” Components	FY05 (4Q)
Software Drop #3 to TIA System	FY06 (1Q)
Evaluate Phase IV Evidential Reasoning Components	FY06 (4Q)
Evaluate Phase IV Collaboration Components	FY06 (4Q)
Evaluate Phase IV “Read Everything” Components	FY06 (4Q)
Software Drop #4 to TIA System	FY07 (1Q)
Evaluate Phase V Evidential Reasoning Components	FY07 (4Q)
Evaluate Phase V Collaboration Components	FY07 (4Q)
Evaluate Phase V “Read Everything” Components	FY07 (3Q)
Software Drop #5 to TIA System	FY07 (4Q)

Wargaming the Asymmetric Environment (WAE)

OVERVIEW: The WAE Program predates both the IAO and the TIA Program. The objective of the WAE Program is to develop automated predictive models “tuned” to the behavior of specific foreign terrorist groups to facilitate the development of more effective force protection and intervention strategies. Specifically, WAE is developing predictive technologies to enable the development of a terrorist-specific continuous indication and warning system that will provide earlier and more specific warnings of future attacks and attack characteristics (target characteristics, tactic, geographical region, timeframe, and adversarial vulnerabilities). Additionally, WAE is developing a terrorist-specific information operations gaming environment to allow decision-makers to better understand their intervention options (deflect, deter, and defeat) through gaming an adversary’s likely future actions and reactions based upon their specific motivations and vulnerabilities. WAE is actively working with both DoD and the Intelligence Community throughout the development, testing, and transition of each of these predictive products.

TECHNICAL APPROACH: WAE’s approach views terrorist behavior in the broader context of its political, cultural, and ideological environment. This predictive modeling approach is an extension of a solid core of behavioral science research that hypothesizes that while individuals and groups may vary the manner in which they execute an attack, their decision to attack is triggered off external events (political, cultural, and ideological) that, in their view, make their action politically advantageous. This differs significantly from the current analytic approach,

which attempts to monitor a group's activity by tracking their planning and logistic functions. Although the two approaches are complimentary, WAE's approach differs from the tracking approach in some significant ways. First, WAE's focus is on select behaviors, such as attack behaviors. The rationale is that WAE is not attempting to establish an overall assessment of a group's capability, but rather to derive the predictive triggers associated with the decision to use that capability. Second, WAE's focus is on deriving predictive patterns from more high-level information associated with the political, cultural, and ideological environment surrounding the group. The rationale for this is the covert nature of group behavior, which by definition attempts to disguise or vary behavior and dilutes most predictive patterns at detailed levels of planning and logistics. Finally, WAE's focus is on deriving triggers that can directly address the question of how the United States can potentially influence the adversary's behavior. The rationale for this is that if groups are triggering off U.S. and Allied political and military behavior, the United States can incorporate these triggers, their own behavior, into a larger information operation campaign designed to deflect, deter, and defeat specific adversaries.

RELATIONSHIP TO TIA: The prototype tools from each phase of WAE will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each WAE tool.

TRANSITION/DEPLOYMENT PLANS: WAE's strategy is to transition the predictive modeling and predictive gaming technologies to DoD and Intelligence operational partners as a part of TIA and in its component form consisting of the continuous indication and warning system and the information operations gaming environment. To date, WAE has, in concert with operational partners, validated several terrorist group specific models against both real-time and historical data. Transition of these predictive technologies and models began in FY 2002 and continues throughout the remainder of the program through FY 2004.

WAE - FY 2004 PRESIDENT'S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$14,836	\$18,604	\$8,221	\$000	FY 2004

PROGRAM SCHEDULE: WAE began in FY 2000 and will conclude in FY 2004.

Milestone	FY/Quarter
Prediction Experiments	FY01/02 (3Q)
Emulation Experiments	FY02 (3Q)
Emulation Experiments	FY02 (4Q)
Generalization Experiments	FY03 (1Q)
Emulation Experiments	FY03 (2Q) - FY04 (4Q)
Prediction Experiments	FY04 (2Q)

Milestone	FY/Quarter
Generalization Experiments	FY03 (4Q)
Emulation Experiments	FY03 (4Q) - FY04 (4Q)
Prediction Experiments	FY04 (2Q)
Generalization Experiments	FY04 (2Q)
Generalization Experiments	FY04 (3Q)
Emulation Experiments	FY03 (3Q)
Test & Transition	FY03 (2Q) - FY04 (4Q)

Rapid Analytical Wargaming (RAW)

OVERVIEW: The objective of the RAW Program is to develop a faster than real-time analytical simulation to support US. readiness for asymmetric and symmetric missions across analytical, operational, and training domains. The program will develop technologies to generate a fuller spectrum of known and emergent behaviors that will provide decision-makers with the ability to better anticipate future political, policy, security, and military/terrorism activity within a region. The operational benefit of RAW includes the ability to monitor key behaviors and actors within a region in real-time and to rapidly game potential U.S./Allied interaction from a political, policy, and military perspective.

TECHNICAL APPROACH: RAW's approach will be to develop and integrate into a single simulation environment: 1) predictive models of countries, key leaders and terrorist groups; 2) analytical decision models; and 3) real-time extraction technology. The predictive models will incorporate the modeling approach and tools from DARPA's WAE Program. The analytical decision models will be developed with our operational partners and will consist of a hybrid of computer-based reasoning technologies. The real-time extraction technology will exploit the extraction work in other DARPA programs. The results will be tested against both real-time and historical data.

RELATIONSHIP TO TIA: The prototype tools from each phase of RAW will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each RAW tool.

TRANSITION/DEPLOYMENT PLANS: If RAW research is successful, the technology will transition as an analytical war game and as a supporting component for TIA. RAW's usefulness will be established based on integration in TIA experimentation, projected to occur no earlier than FY 2005.

RAW - FY 2004 PRESIDENT'S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$000	\$7,500	\$9,360	FY 2007

PROGRAM DURATION: RAW begins in FY 2004 and concludes in FY 2007. A milestone schedule is under consideration.

Futures Markets Applied to Prediction (FutureMAP)

OVERVIEW: The FutureMAP Program provides DoD with market-based techniques for avoiding surprise and predicting future events. Strategic decisions depend upon the accurate evaluation of the likelihood of future events. This analysis often requires independent contributions by experts in a wide variety of fields, with the resulting difficulty of combining the various opinions into one assessment. Market-based techniques provide a tool for producing these assessments. Applications include analysis of political stability in regions of the world, prediction of the timing and impact on national security of emerging technologies, assessment of the outcomes of advanced technology programs, or other future events of interest to DoD. The rapid reaction of markets to knowledge held by only a few participants may provide an early warning system to avoid surprise.

The application of FutureMAP within TIA will answer predictive questions such as “Will terrorists attack Israel with bioweapons in the next year?” To answer this question, FutureMAP would aggregate information from a variety of experts, e.g., analysts for Israel and the Middle East and specialists in bioweapons and other technical areas. The technology question is how to combine this disparate information.

TECHNICAL APPROACH: FutureMAP’s innovation is to use markets to replace today’s approach of discussion and consensus among experts. The new approach is to set up, as it were, a “market” in two kinds of futures contracts: One pays \$1 if an attack takes place; the other pays \$1 if there is no attack. Market participants trade the issued contracts freely. Prices and spreads signal probabilities and confidence. Since markets provide incentives for good judgment and self-selection, the market will effectively aggregate information among knowledgeable participants. This approach has proven successful in predictions concerning elections, monetary policy decisions, and movie box office receipts; DARPA research is investigating its success in Defense-related areas.

DARPA has supported two seedling efforts under the Small Business Innovation Research (SBIR) program to test the feasibility of FutureMAP. One ongoing effort is defining and managing markets to answer specific questions posed by DoD. Typically, these markets will have a small number of invited participants who bring their information together through the market mechanism. We envision markets of 15 to 20 participants addressing questions about the probabilities of specific kinds of failure within our national infrastructure. The results from these markets would be used as input to further analytical steps.

The other ongoing effort is defining and managing long-running markets based on data series that are available from independent news and intelligence sources. The “markets” will allow a wide range of participants to trade futures on composite “securities” that express changes in combinations of series. Composite securities provide a participant, who has insight into interrelations among basic securities, with a means of expressing this insight and benefiting from the expression if correct. A simple composite is an intersection between two basic securities; e.g., the probability that a decrease in Country X gross domestic product will coincide with an increase in Country X civil unrest.

RELATIONSHIP TO TIA: The prototype tools from each phase of FutureMAP will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each FutureMAP tool.

TRANSITION/DEPLOYMENT PLANS: Potential FutureMAP applications within DoD include analysis of political stability in regions of the world, prediction of the timing and impact on national security of emerging technologies, and assessment of the outcomes of advanced technology programs. In addition, the rapid reaction of markets to knowledge held by only a few participants may provide an early warning system to avoid surprise. Interested parties include the Center for Army Analysis and the CIA. FutureMAP predictive technology will be evaluated in a series of TIA experiments at INSCOM beginning in FY 2005. Based on the results of these experiments, successful technology will be transitioned in the form of permanent components of a TIA prototype.

FUTUREMAP - FY 2004 PRESIDENT’S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$000	\$000	\$3,000	\$5,000	FY 2008

PROGRAM SCHEDULE: FutureMAP begins in FY 2004 and concludes in FY 2008. A milestone schedule is under consideration.

Automated Speech and Text Exploitation in Multiple Languages

There are three programs under the Automated Speech and Text Exploitation in Multiple Languages heading:

- Effective, Affordable, Reusable Speech-to-Text (EARS)
- Trans-lingual Information Detection, Extraction and Summarization (TIDES)
- Global Autonomous Language Exploitation (GALE)

Effective, Affordable, Reusable Speech-to-Text (EARS)

OVERVIEW: EARS aims to create effective speech-to-text (automatic transcription) technology for human-human speech, focusing on broadcasts and telephone conversations (the most critical media for a wide range of national security applications) to produce core-enabling technology that can be ported rapidly to many languages and a number of applications.

EARS will drive word error rates down to 5-10 percent (a three-fold reduction from the state of the art for broadcast speech; five-fold for conversations) and extract additional information from the signal. This capability will completely transform the way voice is processed by many organizations: Machines will be able to detect useful material much more accurately; people will be able to read rapidly rather than listen laboriously; and automatic extraction, summarization, and translation of speech will finally become feasible.

Human-human speech is an indispensable source of intelligence. Many organizations within the DoD, the Intelligence Community, and Law Enforcement are charged with monitoring broadcasts or telephone conversations. All are overwhelmed by the magnitude and difficulty of this task. They must confront huge and growing volumes of traffic with fewer people and minimal automation. Foreign languages exacerbate the problem—and are often the only source of vital information.

DARPA believes that EARS could enable a 100-fold improvement in human productivity—10-fold from much more accurate automatic selection and filtering and 10-fold from people reading rapidly instead of listening laboriously. It will enable other software (such as that being developed in the TIDES Program) to populate large knowledge bases with names, entities, and facts extracted automatically from voice signals; to summarize the content of individual telephone calls or sets of related calls; and to provide usable English-language translations of foreign language audio. With EARS, the United States could have 1,000 times more “ears” working on exploiting voice communications than we do now.

TECHNICAL APPROACH: Human-human speech is noticeably different from human-machine speech; converting it to text is much harder. The vocabulary is much larger, the pronunciation is more complex and variable (especially for conversational speech), the speakers are not attempting to be understood by a system, and there is a dearth of human transcripts to learn from (for conversational speech).

EARS is leveraging the impressive achievements of prior DARPA research in speech-to-text technology, recent breakthroughs in speaker identification and statistical natural language processing, a host of promising new technical ideas, huge quantities of speech and text now available electronically, plus very substantial advances in computational power. All of these contribute to the feasibility of attacking and conquering the EARS challenges.

The basic approach is to treat speech production as a stochastic encoding process and to cast speech-to-text as decoding in a probabilistic framework. The underlying acoustic and language models are being radically revised to exploit information known about human articulatory,

auditory, and linguistic processes. Parameter values will be automatically learned from huge quantities of data. Metadata (information about speakers, topics, names, new words, structure, emphasis, and emotion) will be automatically extracted, fed back to improve the transcription process, and fed forward to be part of the output—measurably enriching the stream of words and making the output maximally useful to both people and machines.

The core algorithms will be adapted to two media (broadcast news and telephone conversations) and three languages (English, Chinese, and Arabic). Algorithms will be formally evaluated for accuracy at 12-month intervals using procedures designed and administered by National Institute of Standards and Technology (NIST).

Teams of experienced researchers from leading academic and industrial research laboratories are doing the research. Each team is investigating a wide range of promising ideas and will integrate the most successful ones into that team's evolving system.

To facilitate research and evaluation, broadcasts and telephone conversations are being collected and annotated. DARPA affirms that the broadcasts collected were produced for public consumption and are being acquired in accordance with copyright restrictions; the conversations are from volunteers who are paid for the right to use their speech.

RELATIONSHIP TO TIA: The prototype tools from each phase of EARS will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each EARS tool.

TRANSITION/DEPLOYMENT PLANS: As it matures, EARS technology will be tried in various TIA experiments and in non-TIA experiments at several other agencies, most notably the National Security Agency (NSA). The first transitions are likely to occur in 2004 and to continue incrementally for several years, as EARS accuracy, richness, and robustness improves.

EARS technology will fit easily into existing systems and efforts that employ speech-to-text. These include research initiatives (e.g., Translingual Information Detection, Extraction and Summarization [TIDES] being developed by DARPA) and operational capabilities (e.g., Open Audio Source Information System [OASIS] used by the Foreign Broadcast Information Service) that employ but do not create state-of-the-art speech-to-text for human-human communication in one or more languages.

EARS will facilitate the development of techniques for exploiting speech in multispeaker environments (e.g., command centers, teleconferences, and meetings); enable machines to monitor discussions among people and proactively bring important information to their attention; search and mine vast audio archives; and produce timely transcripts for rapid reading, dissemination, and reaction.

Most significantly, EARS will enable a large number of revolutionary new applications that require higher accuracies. These include rapid reading instead of laborious listening, precision targeting (spotting) of key conversations, automatic translation of foreign language speech transcripts, and automatic population of large knowledge bases using information extracted from volumes of human-human communications.

Rapid reading will be comparatively easy to transition, because it is easy to implement. DARPA will work with its partners in the military and Intelligence Community to ensure this transition/transformation happens.

Precision targeting will be slightly more challenging, but is of great interest to high-volume customers like NSA, which are struggling with the twin challenges of surging volumes and dwindling staff.

Automatic population of large knowledge bases from text is actively being discussed within the Intelligence Community. EARS will enable this work to be extended to include audio sources that contain vital information that would otherwise, as a practical matter, lie out of reach.

PROGRAM SCHEDULE: The EARS Program began in FY 2002 and will conclude in FY 2007.

Milestone	FY/Quarter
Evaluate performance at end of Phase I	FY03 (4Q)
Evaluate performance at end of Phase II	FY04 (4Q)
Evaluate performance at end of Phase III	FY05 (4Q)
Evaluate performance at end of Phase IV	FY07 (2Q)
Demonstrate porting to a new language in 1 month	FY06 (2Q)
Demonstrate porting to a new language in 1 week	FY07 (2Q)

Translingual Information Detection, Extraction, and Summarization (TIDES)

OVERVIEW: TIDES aims to make it possible for English speakers to find and interpret needed information quickly and effectively, regardless of language or medium. Source data could be unformatted raw audio or text, stationary or streaming. Critical information could span one or more documents, one or more places, and one or more languages.

To create that capability, TIDES is developing a suite of component technologies, integrating those components to maximum effect in technology demonstration systems, and experimenting with the systems on real-world problems. These are all high-risk research activities.

The component technologies fall into the following classes:

- Detection - Find or discover information needed by an operator/analyst.
- Extraction - Extract key information about entities, relations, and events.

- Summarization - Substantially reduce the amount of material that an operator/analyst must read.
- Translation - Convert raw language text, audio transcripts, or summaries into English.

Detection, extraction, and summarization must work both within and across languages; translation must work from other languages into English. In addition to creating effective technology, TIDES aims to develop methods for porting these technologies rapidly and inexpensively to other languages, including those having severely limited linguistic resources.

TIDES is integrating the component technologies with one another and with other technologies to produce synergistic, effective, end-to-end technology demonstration systems able to address multiple operational needs. The goal is not simply to increase the productivity of operators and analysts, but also to provide commanders and other decision-makers with a great deal of vital information that, as a practical matter, is currently out of reach.

The TIDES effort has the potential to address a significant national security issue. U.S. forces must be able to operate around the globe, often on short notice, in regions where English is not the native language. To be effective, and to protect themselves, our forces must be able to understand a wide variety of information that is available only in foreign languages and to know what is being said in a region by and to the local populace. There are about 228 countries whose people speak approximately 6,700 languages. DoD is currently interested in about 200 languages, and the list constantly changes. Military and civilian analysts and translators with suitable foreign language skills are in short supply, slow to train, and difficult to retain.

TIDES will mitigate all these problems by enabling English-speaking operators and analysts to find and interpret relevant foreign language information.

TECHNICAL APPROACH: The key technical challenge for TIDES is the development of translational technology that is sufficiently robust and accurate to be a real force multiplier. Even monolingual technology is hard, and everything becomes more difficult when linguistic resources (e.g., annotated speech and text, lexicons, and grammars) are scarce. TIDES is pushing the envelope in all these areas.

Most of the research is being conducted in three key languages: English, Chinese, and Arabic. Stress tests are conducted on surprise languages to ensure the portability of the technology.

TIDES is leveraging a great deal of successful work in prior DARPA programs plus promising research going on around the world, especially new work on statistical natural language processing. TIDES will take advantage of significantly enhanced computational power; and it will exploit the rapidly growing volumes of speech and text accessible electronically, including parallel text.

The most productive solutions are expected to be combinations of techniques, both statistical and symbolic. To the extent to which algorithms can be made to learn from lightly annotated data, it will be possible to make TIDES technology more robust and more rapidly and inexpensively portable to new languages that suddenly become operationally important.

To detect information specified by a user, researchers are employing probabilistic vector state models enhanced with query expansion techniques, event-situated named entities, and multiple bilingual term translations. To discover potentially useful new information, researchers are using topic-conditional models plus finer-grained models, comparing incoming data to previously identified events.

To extract key information about entities and relationships, researchers are developing active learning techniques that take advantage of manually annotated data (tag-a-little, learn-a-little), linguistic pattern discovery techniques able to exploit large unannotated corpora in multiple languages, plus a variety of statistical pattern recognition models. The extracted information will aid detection and summarization.

To summarize the content of one or more documents or automatically transcribed audio segments, researchers are developing automatic headline generation techniques using hidden Markov models plus extractive and concatenative synthesis techniques that reassemble the key information in a logical order. These techniques will be able to “learn” from examples of humanly generated summaries.

For translation, researchers are investigating example-based and statistical translation approaches that exploit the increasing availability of parallel text resources. Example-based translation looks for matching fragments and patterns of text, then reassembles them. Statistical translation approaches model foreign language input as if it were a corrupted version of English, then seeks to recover the “original” English, finding the signal buried in the noise.

In all areas, researchers will develop and use techniques to construct bilingual dictionaries automatically from parallel or comparable corpora and to learn grammar rules automatically from tagged and bracketed text known as *treebanks*.

Meaningful objective performance measures are being used, including a novel automated method for evaluating translation quality that is greatly accelerating progress. NIST will oversee all the formal evaluations.

RELATIONSHIP TO TIA: The prototype tools from each phase of TIDES will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each TIDES tool.

TRANSITION/DEPLOYMENT PLANS: TIDES technology is being evaluated in various TIA experiments and in non-TIA experiments at several other agencies, most notably the CIA. These evaluations have been in progress for several years.

During the past 2 years, TIDES has combined various detection, extraction, summarization, and translation technologies into several text and audio processing (TAP) systems: MiTAP, OnTAP, and ViTAP. TIDES is now producing relatively robust, reconfigurable technology components for use in TIA, in a new unified TAP system, and for possible transition to other agencies.

TIDES technology has been employed in all TIA experiments and in a series of TIDES-specific, user-centric integrated feasibility experiments (IFEs). Each experiment has helped us assess and refine the technology and will facilitate the transfer, when the technology works well enough, into operational military and intelligence systems.

PROGRAM SCHEDULE: The TIDES Program began in FY 2002 and will conclude in FY 2005.

Milestone	FY/Quarter
Component Technology Research	
Select principal focus languages.	FY01 (2Q)
Define clear research objectives for component technologies.	FY01 (4Q)
Conduct baseline evaluations of detection, extraction, and summarization.	FY01 (4Q)
Demonstrate monolingual detection, extraction, and summarization.	FY02 (1Q)
Demonstrate enhanced translingual detection.	FY03 (1Q)
Conduct baseline evaluation of translation.	FY02 (2Q)
Demonstrate enhanced translation capability.	FY03 (2Q)
Demonstrate initial translation capability for new language in 3 months.	FY04 (1Q)
Demonstrate enhanced translation capability for new language in 1 month.	FY05 (1Q)
Technology Integration and Experimentation	
Assemble MiTAP System.	FY01 (3Q)
Conduct IFE-Bio-1.	FY01 (3Q)
Assemble OnTAP System.	FY01 (3Q)
Conduct IFE-Bio-2.	FY02 (2Q)
Conduct IFE-Arabic-1.	FY02 (3Q)
Conduct IFE-Translingual-1.	FY03 (3Q)
Conduct IFE-Translingual-2.	FY04 (3Q)

Global Autonomous Language Exploitation (GALE)

OVERVIEW: GALE aims to make it possible for machines to discover critical foreign intelligence information in a sea of human language (speech and text) from around the globe, delivering it in actionable form to military operators and intelligence analysts without requiring them to issue specific requests.

The intent is to greatly enhance the timeliness and completeness of intelligence production by exploiting large volumes of heterogeneous material autonomously, thereby magnifying the impact of the skilled operators and analysts who are overwhelmed and in dangerously short supply.

If GALE succeeds, machines will be able to find, refine, combine, and package information from broadcasts, conversations, newswire, and Internet sources; discover trends and deviations; discern operator/analyst interest from their actions and reports; and issue critical alerts, reports, and pointers whenever appropriate without overwhelming the operator/analyst whom they serve.

TECHNICAL APPROACH: GALE will build off the essential groundwork being laid by TIDES and EARS; improve it as needed; and exploit recent advances in machine learning, intelligent alerting, and database technology.

GALE will exploit raw language data (not structured information like other programs), automatically populate a knowledge base with metadata and associations derived from the speech and text, and proactively determine the particular information that a particular operator/analyst should see.

The outputs of GALE would be combined with the outputs of other ongoing or anticipated programs that exploit structured data.

RELATIONSHIP TO TIA: The prototype tools from each phase of GALE will be passed to the TIA Program for experimentation and evaluation. The feedback from TIA experimentation and evaluation will be used to guide subsequent development of each GALE tool.

TRANSITION/DEPLOYMENT PLANS: GALE is proposed as an FY 2004 new start. It will make its first transitions (via TIA and non-TIA experiments) in 2005. GALE technology will be rapidly refined in response to customer feedback. The principal customers outside of DARPA are DIA, CIA, and NSA.

GALE is another key step toward DARPA's vitally important goal of teaching computers to "hear," "read," and "understand" human language in all its forms.

PROGRAM SCHEDULE: The GALE Program begins in FY 2004 and concludes in FY 2009. A milestone schedule is under consideration.

GALE - FY 2004 PRESIDENT'S BUDGET (\$000) - EARS, TIDES and GALE:

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$27,831	\$34,174	\$46,332	\$48,383	FY 2009

Situation Presentation and Interaction

There are two programs under the Situation Presentation and Interaction heading:

- Babylon
- Symphony

Babylon

OVERVIEW: Babylon predates both the IAO and the TIA Program. Babylon is not planned for integration in TIA. However, information on Babylon is provided for completeness of this report since all natural language processing programs have been concentrated in this office. The goal of the Babylon Program is the development of natural language two-way translation technology to support military field operations and other agencies requiring real-time field-oriented translation support.

TECHNICAL APPROACH: Efforts are divided into four major task categories:

- **DARPA 1+1:** This task is creating a limited-use handheld translation device as a replacement for the aging Phraselator, DARPA's original technology (formerly known as the DARPA One-Way). The 1+1 technology is centered on the use of highly constrained dialog phrases for the English speaker based on the desired activity (e.g., checkpoint activities, medical first response, or refugee support) with natural language translation for the foreign speaker tied to the phrase used by the English speaker. By retaining the constraints on the English speaker, the ability to adapt response models for the foreign speaker is made easier to support rapid development and delivery. Development efforts under the 1+1 include the translation software (Pashto) and algorithms for insertion into the next-generation handheld translation device; alternative translation software (Arabic) and algorithms for insertion into the next-generation handheld device; and the next-generation handheld device itself, which will be the platform for the 1+1 and some Two-Way (see below) software packages.
- **DARPA Two-Way:** This is a basic research effort to develop domain-constrained natural language multilingual dialog systems so both English and foreign speakers may use full natural language. The use of phrases is completely eliminated in this research. The users are constrained to specific domains, e.g., force protection, medical triage/first response, refugee support, maritime intercept, and other operational tasks as required by specific users. The Two-Way technologies are being developed by multiple teams using a variety of methods and algorithms, each competing to be declared superior for future development. Development efforts under the Two-Way include scalable translation software (Mandarin Chinese/American English) and algorithms for insertion into multiple platforms for use in force protection and medical domains, translation software (Pashto/American English) and algorithms for insertion into the next-generation handheld device for use in force protection and medical domains, translation software (Dari/Farsi/American

English) and algorithms for insertion into the Army's Land Warrior platform for use in force protection and medical domains, and a wearable platform-based system focused on Pacific Rim languages.

- **Data Collection and Evaluation:** To develop and evaluate new translation technologies in the domains required by the DoD and other agencies, new data collections and evaluation protocols must be developed. This task supports all necessary infrastructure (equipment, personnel, access coordination, evaluation coordination and execution) for development and evaluation of new systems. In addition, to support future sustainment for translation technologies, a language center repository is necessary to serve as a central point of support for DoD systems after the Babylon program has ended.
- **International and Coalition Collaborative Research:** To ensure that its research and technology deliverables remain the finest in the world, DARPA actively seeks collaborative relationships with the best international research teams. DARPA also seeks collaboration with its counterparts in coalition defense organizations. Babylon has entered into a collaborative research agreement with the European Union's AMITIES Program. This effort is developing multilingual dialog systems supporting kiosk and phone center operations. The intellectual exchange between all the members has advanced the state of the art for all participants. DARPA funding is limited to U.S. performers (as European funding is limited to EU members). Coalition support is just starting the negotiation process.

RELATIONSHIP TO TIA: There are no plans for integration with TIA.

TRANSITION/DEPLOYMENT PLANS: Babylon technology is transitioning to support operational systems in the Navy and Marine Corps. It is also planned for support of DoD and other agency operational requirements for force protection, medical triage/first response, refugee support, maritime intercept, and other operational tasks as required by specific users. The program is participating in demonstrations, field experiments, and exercises as part of the Language and Speech Exploitation Resources (LASER) Advanced Concept Technology Demonstration (ACTD). Forums for testing and evaluating the proposed technology will include planned exercises, demonstrations, and real-world events. A series of small-scale military utility assessments (MUAs) of tactics, techniques and procedures (TTP) and equipment inserted into already planned exercises will be ongoing during the first 3 years of the ACTD. The program is also evaluating technology transition to European Union members via the AMITIES Program, initially to support the Royal Marines (UK) English-for-Pashto and Arabic translation requirements.

PROGRAM SCHEDULE: The Babylon Program began in FY 2002 and will conclude in FY 2004.

Milestone	FY/Quarter
DARPA 1+1	
Language conversion to prototype platforms	FY02 (4Q)
Interface modifications for limited two-way (1+1) support	FY02 (4Q)
Board modifications and upgrade to X-scale CPU	FY03 (1Q)
Hardware patches and fixes	FY03 (2Q)
Production run complete, 200 New X-scale systems	FY03 (2Q)
Initial system delivery	FY03 (3Q)
End user Training completed	FY03 (4Q)
Babylon Teams	
Two-way language development (includes date collection and corpora development)	FY02 (4Q)
Two-way interface development	FY02 (4Q)
Core Interlingua development	FY03 (2Q)
Shallow parser development	FY02 (4Q)
Translation integration	
Large (Pentium CPU and LandWarrior)	FY03 (3Q)
Small (PDA)	FY03 (3Q)
ASR optimization for multiple languages	FY02 (4Q)
Data Collection and Evaluation	
Research languages and domains selected	FY02 (3Q)
SMEs assigned to research teams	FY02 (4Q)
Language collection and delivery to teams	FY03 (2Q)
Collection development complete (development, training, and evaluation sets)	FY03 (3Q)

Milestone	FY/Quarter
Dry-run evaluation on prototypes	FY03 (4Q)
Evaluation refinement and metrics validation	FY04 (1Q)
Formal Babylon evaluation (validated base establishment)	FY04 (4Q)
Publication of evaluation and metric for multilingual speech-to-speech translators	FY04 (4Q)

Symphony

OVERVIEW: The Symphony Program is a follow-on to DARPA’s Communicator Program. Symphony is targeted at the development of natural language dialog technology to support military field operations and other agencies requiring real-time, field-oriented dialog systems. These systems are oriented to such operational tasks as ordering logistics, coordinating calls for fire support, and friendly passage of lines. Symphony is not planned for integration in TIA at this time. However, information on Symphony is provided for completeness of this report.

TECHNICAL APPROACH: Program efforts are divided into four major task categories:

- **Applied Systems Development:** Development teams will be tasked to build a dialog system for a selected organization having an immediate need in an operational environment. Examples of these operational dialog systems include a shipboard dialog system that provides a ship’s status to an authorized intercom user; the Army’s Battlefield Casualty Reporting System (BCRS), which will automate the current paper intensive (and slow) process of getting casualty information back for family notification and strength reporting; an aircraft maintenance mentor dialog system; and a navigational dialog system designed to support real-time vehicle navigation in complex urban terrain.
- **Core Research for Dialog Technology:** While the developmental teams will continue to research and develop dialog components, a targeted core research effort is required to ensure the generalizability of the dialog architecture beyond the environments evaluated in Communicator, the original program that developed a generalized dialog architecture called Galaxy. The areas of research include methods and algorithms to ensure the effectiveness of dialog systems in meeting and multispeaker environments, the use of prosodics and mixed initiative in the dialog management system, and optimization of automatic speech recognition (ASR) for dialog systems in noisy environments.
- **Data Collection and Evaluation:** To support the development teams and to evaluate the effectiveness of each implementation, a solid set of evaluation protocols will be required. In addition, an integrator will be required to maintain the Galaxy Architecture and execute evaluations. The tasks required within this category involve

planning and execution of the domain-independent evaluations for the systems developed under the applied research task, development of the domain-independent evaluation protocols, maintenance and upgrade of the Galaxy Architecture or its replacement, and evaluation of the domain-specific systems as they perform in their designed task.

- **International and Coalition Collaborative Research:** To ensure that our research and technology deliverables remain the finest in the world, DARPA actively seeks collaborative relationships with the best international research teams. DARPA also seeks collaboration with its counterparts in coalition defense organizations to ensure interoperability and smooth integration into coalition operational plans.

RELATIONSHIP TO TIA: There are no plans for integration with TIA at this time.

TRANSITION/DEPLOYMENT PLANS: Support to military field operations as well as other agencies requiring real-time field-oriented dialog systems to support operational tasks such as ordering logistics, coordinating calls for fire support, and friendly passage of lines. Development teams are responsible for building a dialog system for a selected organization having an immediate need in an operational environment. Examples of these operational dialog systems include a shipboard dialog system to provide ship's status to an authorized intercom user, Army BCERS to automate the process of getting casualty information back for family notification and strength reporting, an aircraft maintenance mentor dialog system, and a navigational dialog system to support real-time vehicle navigation in complex urban terrain.

PROGRAM SCHEDULE: The Symphony Program is planned to begin in FY 2004 and conclude in FY 2006. A milestone schedule is under consideration.

BABYLON AND SYMPHONY - FY 2004 PRESIDENT'S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$15,901	\$8,770	\$10,869	\$7,500	FY 2006

**Bio-Event Advanced Leading Indicator Recognition Technology
(Bio-ALIRT)**

OVERVIEW: The Bio-ALIRT Program predates both the IAO and the TIA Program. The objective of the Bio-ALIRT Program is to develop technology for early (i.e., prior to when people begin to seek professional medical care) detection of a covert biological attack. Earlier detection enables earlier intervention and may enable drastic reductions in fatalities as well as better use of scarce public health resources.

Bio-ALIRT will analyze nontraditional data sources (i.e., aggregate and anonymized data about human behaviors and about sentinel animals), correlating these sources with known natural outbreaks of disease, to determine which data sources provide the earliest and most specific leading indicator of an outbreak. Example data sources include numbers of school or workplace

absences, number of calls to poison control centers or nurse hot lines, and purchases of over-the-counter (OTC) pharmaceuticals. Flu outbreaks can be used as a surrogate for the types of pathogens that might be used by a rogue state or terrorist because the early symptoms are indistinguishable from them. Bio-ALIRT is collecting and analyzing these data for several U.S. cities, including the National Capital Area and the Hampton Roads area of Virginia, because attacks in these cities would result in a degradation of military leadership or deployment capabilities.

An extremely beneficial side effect of the Bio-ALIRT research is the current monitoring capability of the National Capital Area for potential outbreaks. DARPA affirms that all Bio-ALIRT data is obtained in accordance with the Health Insurance Portability and Accountability Act (HIPAA) and other privacy requirements. To provide even more than the required level of privacy protections, Bio-ALIRT is also conducting research to develop techniques and algorithms to provide formal, provable assurance that aggregate and anonymous data cannot be re-identified.

Bio-ALIRT results about what data sources are most useful and what algorithms provide the best detection capability from these data sources can be used not only for military force protection systems, but also for homeland security defense against biological attacks. Program plans include making this knowledge available to the Departments of Homeland Security and the Department of Health and Human Services for their use.

TECHNICAL APPROACH: Bio-ALIRT analyzes these existing and authorized electronic data streams and looks for spikes in gross numbers (e.g., numbers of school or workplace absences, number of calls to poison control centers or nurse hot lines, and purchases of OTC pharmaceuticals) that may signify a disease outbreak, rather than performing data mining or seeking any individual information. If signs of an outbreak are detected, the system alerts a certified public health official authorized to conduct epidemiological research into outbreaks. That doctor would then have the opportunity to follow up with local medical providers by inquiring into cases or recommending that local providers perform diagnostic tests on patients to differentiate terrorist-type disease from normal flu-like illness. It would be these early confirmatory tests that might trigger a full-blown and timely public health response rather than the Bio-ALIRT technology itself. If Bio-ALIRT has a false alarm, it may cause additional work for a local public health official, but it will not be a public event.

Technical challenges in the Bio-ALIRT Program include determining the value of each data source, alone and in combination with others, for earlier outbreak detection; correlating/integrating information derived from heterogeneous data sources; development of autonomous signal detection algorithms with high sensitivity and low false alarms; creation of disease models for autonomous detection; and maintaining privacy protection while correlating depersonalized data sources.

There are four Bio-ALIRT development projects plus an additional one that provides evaluation. The four projects are identifying and developing both nontraditional data sources such as OTC medications, 9-1-1 emergency calls, utility usage, cough detectors on a military base, and even animal health data and measuring how they correlate to, and anticipate in-time, “gold standard” medical data that shows historically when flu-like illnesses appear in a community. Our belief is

that in a terrorist outbreak, such behaviors will also be a leading indicator of the outbreak of anthrax or other pathogen of interest. However, due to the “nonstationarity” of the data, one cannot use a simple autoregressive moving average to capture that outbreak, but rather must factor in day-of-the-week effects, seasonal effects, promotional effects, etc.

Two of the Bio-ALIRT development projects are large-scale prototype contracts and two are technology development contracts. The geographic and temporal variance of the data sources being examined requires that potential results be evaluated at a realistic scale and with realistic characteristics. The two prototype contractors are developing prototypes that process nontraditional and “gold standard” data in a geographic area, applying their detection algorithms to identify disease outbreaks. Bio-ALIRT does not seek to develop operational production systems, per se, but needs to have functional prototypes for realistic operation and evaluation to show the value of the algorithms and data sources in a realistic environment.

The two technology contractors develop new algorithms, simulations, and novel data sources to test and provide to the systems contractors. Quantitative evaluations of algorithms and of data sources occur annually. They have shown how they have related their anonymized and aggregated data sources against historical outbreaks of flu-like illness. In addition, their developmental algorithms have been tested against a simulated outbreak generated by a computer model. This process was supervised by the fifth (evaluation) contractor.

RELATIONSHIP TO TIA: Bio-ALIRT technology is intended to trigger an earlier response to one form of asymmetric attacks that have already occurred versus early identification and preemption of planned attacks. As such, there are no definitive plans to directly integrate Bio-ALIRT prototype systems in TIA, although there is significant potential for software agents and algorithms developed in Bio-ALIRT to be applied in support of TIA nonbiosurveillance requirements. For this reason, R&D efforts in Bio-ALIRT are coordinated with system development activities in TIA.

TRANSITION/DEPLOYMENT PLANS: The Bio-ALIRT Program has technologies that are being implemented and transitioned. Bio-ALIRT is helping provide technical support for the surveillance of the medical well-being of all nondeployed U.S. forces worldwide in cooperation with the DoD Global Emerging Infections System (GEIS) ESSENCE program, which monitors standard ambulatory data records in accordance with prescribed procedures. GEIS ESSENCE has detected two outbreaks that were previously unknown to local military medical authorities. In addition, Bio-ALIRT is cooperating with GEIS to detect outbreaks of disease in the National Capital Area by leveraging anonymized and/or aggregated data in the cooperative ESSENCE II project. A prototype capability will be ready to transition to Maryland in 2003. A final prototype will be ready to transition to the military by the end of FY 2004.

Another Bio-ALIRT effort has an advanced detection algorithm that is downloadable by health departments around the country and was used at the Salt Lake City Olympics as part of the Real-time Outbreak Detection System (RODS). The RODS, which was developed outside of DARPA, is still in use in Utah and Pennsylvania. This supporting algorithm development effort will have an advanced anomaly detection and spatial scan statistic as an option for integration into RODS and deployment by the end of FY 2003. Advanced versions of the detection algorithms will be available for use by public health departments, including the military, in 2004.

A third Bio-ALIRT project has provided its statistical anomaly detector to the Naval Medical Center (NMC)-Portsmouth, where it is in use by on-site Navy staff for early detection of outbreaks. A memorandum of understanding (MOU) is in place that provides for the eventual sharing of military medical information with the Virginia Department of Public Health, which Bio-ALIRT also hopes to support with anonymized and aggregate information in the coming year (2003). DARPA affirms that any sharing of information under this MOU will be in accordance with all applicable laws. Bio-ALIRT will have software components to transition to operational use at NMC-Portsmouth and Air Force and Army hospitals in Hampton Roads, as well as rapid population health detector components ready for use within city public health departments and private sector medical facilities in Hampton, Virginia. They will have a final integrated prototype package ready for deployment at the end of FY 2004.

The fourth Bio-ALIRT project is looking at active surveillance techniques, where data is provided voluntarily, for more permissive environments such as military bases. The contractor has received internal institutional approval to conduct surveys of employees at its campus to try to detect outbreaks of flu-like illness. A car-counting software package to detect differences in road traffic is available for download. Environmental background data against which to measure the spread of disease is available, as is a privacy protection algorithm based on the k-anonymity technique developed by Carnegie Mellon University.

The Bio-ALIRT biosurveillance program is making significant technical progress and holds significant promise to dramatically increase DoD's ability to detect a clandestine biological attack up to 2 days earlier using existing data sources, in time to respond effectively and avoid potentially thousands of casualties. It may also help to meet the military's established requirements for biosurveillance of its own ranks for Force Protection.

BIO-ALIRT - FY 2004 PRESIDENT'S BUDGET (\$000):

<u>FY 2002</u>	<u>FY 2003</u>	<u>FY 2004</u>	<u>FY 2005</u>	<u>Completion Date</u>
\$12,920	\$14,173	\$6,276	\$000	FY 2004

PROGRAM SCHEDULE: Bio-ALIRT began in FY 2001 and will conclude in FY 2004.

Milestone	FY/Quarter
Compile a set of critical BW agents and threat scenarios.	FY02 (2Q)
Construct an archive of historical epidemiological data for normal diseases.	FY02 (3Q)
Complete development of a software environment to emulate a biosurveillance system.	FY03 (1Q)
Complete development of epidemiological models of normal diseases for use by the signal detection algorithms.	FY03 (4Q)

Milestone	FY/Quarter
Detect previously unknown natural disease outbreak(s).	FY02 (4Q)
Determine measurement and performance requirements for various components of the biosurveillance system.	FY03 (2Q)
Complete development of initial signal detection algorithms.	FY03 (4Q)
Complete development of privacy protecting architecture for the integration of heterogeneous data systems .	FY03 (4Q)
Detect disease outbreak 1 day faster than FY 2002 baseline.	FY03 (4Q)
Complete integration of prototype biosurveillance system.	FY04 (2Q)
Field experiments using prototype biosurveillance system.	FY04 (4Q)
Detect disease outbreak 1 day faster than FY 2003 baseline.	FY04 (4Q)

Appendix C – Information Paper on Intelligence Oversight of INSCOM’s Information Operations Center (IOC)

IAJA

15 January 2003

INFORMATION PAPER

SUBJECT: Intelligence Oversight of INSCOM’s Information Operations Center (IOC)

1. Executive Order 12333, signed 4 December 1981 by President Reagan, gives the Intelligence Community its authority to collect foreign and domestic intelligence and counterintelligence information. Within DoD, the EO is implemented by DoD Regulation 5240.1-R; within Army, it is implemented by AR 381-10. For Signals Intelligence information, the EO is implemented by USSID 18.
2. Information that identifies a U.S. person may be collected by a DoD intelligence component only if it is necessary to the conduct of a function assigned the collecting component and falls within one of thirteen specified categories (See DoD Regulation 5240.1-R). In all instances, U.S. person information related to the mission of INSCOM (foreign intelligence and counterintelligence, including international terrorism) may be collected, retained, and disseminated to appropriate authorities. If U.S. person information not within INSCOM's mission is received (through liaison, data mining, etc.), it is either passed to an appropriate agency or purged from the system.
3. Within the Signals Intelligence Community, U.S.SID 18 similarly restricts collection of U.S. person information. As stated in *U.S. IDENTITIES IN SIGINT (U)*, "In the decision process, the protection of the right to privacy of the U.S. person must be weighed against the need of the government to produce foreign intelligence." Within the IOC, USSID 18 controls are built into databases to minimize (i.e., filter out) U.S. person information. In a system that literally collects hundreds of millions of events every day, inadvertent collection of U.S. person information does occur. An analyst won't know, however, that inadvertently collected U.S. person information resides in a database until he or she does a search for lawful mission-related information. In the event U.S. person information is relevant to the mission and should be included in an intelligence product report, such information must receive a legal and intelligence oversight review prior to publication. Depending on the nature and importance of the information, the U.S. person's identifying data

IAJA

SUBJECT: Intelligence Oversight of INSCOM's Information Operations Center (IOC)

may be permitted to stand "as is," may be changed to more generic terms, or may not be allowed at all.

4. The methodologies used to extract information from intelligence databases do not affect the intelligence communities' requirement, at the outset, to collect only that U.S. person information that it may lawfully collect. The hardware and software tools being developed in conjunction with DARPA to mine existing databases do not give us access to any information to which the Intelligence Community does not already have lawful access. It is hoped that these tools once developed will allow an analyst to more rapidly query many disparate databases in order to give real time indications and warnings of terrorist activity.

5. In summary, there are stringent controls in place to ensure no unauthorized U.S. person information is incorporated into INSCOM intelligence products. The software tools being developed in partnership with DARPA do not give us access to information not already in our lawful possession but, hopefully, will speed up the process by which relevant information is extracted, analyzed and used to prevent terrorist activity. In the event there is unlawful collection, retention, or dissemination of U.S. person information in an INSCOM intelligence product report, the violation will be thoroughly investigated and reported to HQDA (DAIG-IO).

COL Schmidli/2555

Appendix D – TIA Program Directives

The Director of DARPA's Information Awareness Office (IAO) issued Terrorism Information Awareness (TIA) Program Directives to ensure that Government program managers and performing contractors involved in TIA experimentation fully understood privacy policies and regulations, the need to protect intelligence sources and methods, and to formally document responsibilities in this context with regard to day-to-day program execution.

Directives contained in this appendix:

- TIA Program Directive Number 1, Intelligence Oversight Training
- TIA Program Directive Number 2, Data Containing Information About U.S. Persons
- TIA Program Directive Number 3, Use of Synthetic Data
- TIA Program Directive Number 4, Memoranda of Agreement with Partners in Experimentation
- TIA Program Directive Number 5, Resources Provided by Other Government Agencies

07 APR 03

Subject: Total Information Awareness (TIA) Program Directive Number 1, Intelligence Oversight Training

From: TIA Program Manager

To: All TIA Contract and Government Personnel Involved with TIA Test Nodes

1. Purpose: The purpose of this TIA Program Directive is to establish an immediate and continuing requirement for all contract and government personnel associated with TIA program test nodes to receive annual intelligence oversight training.

2. Background: The goal of the Total Information Awareness (TIA) program is to significantly increase the ability of the United States to detect, classify and identify foreign terrorists – and decipher their plans – and thereby enable our nation to take timely action to successfully preempt and defeat terrorist acts. To support this goal the Defense Advanced Research Projects Agency's (DARPA) Information Awareness Office (IAO) developed TIA as a multi-year program consisting of the iterative development, acquisition, testing, refinement and integration of advanced technology and processes. A determination was made to use DoD intelligence entities as test nodes. Other nodes within the Intelligence Community may be established. The US contract support personnel is integral to the TIA program and indeed the program largely consists of contractor personnel executing the program under DARPA guidance and leadership.

3. Program Guidance:

a. The TIA program will operate within all applicable laws, executive orders and departmental regulations.

b. All program associated plans and activities conducted by anyone associated with or operating in support of TIA will conform to all applicable laws, executive orders and departmental regulations.

c. TIA is a research and development activity that for experimentation, development, and demonstration purposes operates within DoD and U.S.G intelligence activities. As such, TIA activities, actions and personnel are subject to all applicable U.S.G and DoD Intelligence Oversight rules, regulations, policies and procedures.

d. Effective immediately, all persons associated with TIA and TIA-related programs who perform duties in direct support of TIA test nodes, to specifically include DARPA, contract and contract support personnel, will receive annual Intelligence Oversight training as provided by the test node to which they are assigned. DARPA, contractor, and contract support personnel are not involved in intelligence collection. These personnel are onsite at the various test nodes to assess the usefulness of the TIA technologies and provide technical assistance to the operational users who are participating in TIA testing. DARPA and its contractor personnel who support TIA test nodes are required to receive annual intelligence oversight training to make them aware of the sensitivity of the data that is being processed in the TIA test environment. Within

30 days of the effective date of this Program Directive, all current TIA and TIA associated personnel who support TIA test nodes will receive Intelligence Oversight training. All personnel who become associated with TIA test nodes after the effective date of this letter must receive intelligence oversight training within 30 days of association with the program. Compliance with this directive will be tracked in accordance with the guidance provided in the implementing instructions.

4. Implementing Instructions:

- a. My deputy, Dr. Robert Popp, will administer this Intelligence Oversight program.
- b. Implementing instructions in support of this program directive will be issued under separate cover.

A handwritten signature in black ink, appearing to read "John Poindexter". The signature is fluid and cursive, with a long horizontal stroke at the end.

Dr. John M. Poindexter
Director
Information Awareness Office

07 APR 03

Subject: Total Information Awareness (TIA) Program Directive Number 2,
Data Containing Information About U.S. Persons

From: TIA Program Manager

To: All TIA Contract and Government Personnel Involved with TIA

1. Purpose: The purpose of this TIA Program Directive is to establish an immediate and continuing policy concerning the acquisition or use of data that contains information about U.S. persons.

2. Background: The goal of the Total Information Awareness (TIA) program is to significantly increase the ability of the United States to detect, classify and identify foreign terrorists – and decipher their plans – and thereby enable our nation to take timely action to successfully preempt and defeat terrorist acts. To support this goal the Defense Advanced Research Projects Agency's (DARPA) Information Awareness Office (IAO) developed TIA as a multi-year program consisting of the iterative development, acquisition, testing, refinement and integration of advanced technology and processes. A determination was made to use DoD intelligence entities as test nodes. Other nodes within the Intelligence Community may be established. The contract support personnel is integral to the TIA program and indeed the program largely consists of contractor personnel executing the program under DARPA guidance and leadership.

3. Program Guidance:

a. The TIA program will operate within all applicable laws, executive orders and departmental regulations.

b. All program associated plans and activities conducted by anyone associated with or operating in support of TIA will conform to all applicable laws, executive orders and departmental regulations.

c. TIA is a research and development activity that for experimentation, development, and demonstration purposes operates within DoD and U.S.G intelligence activities. As such, TIA activities, actions and personnel are subject to all applicable U.S.G and DoD Intelligence Oversight rules, regulations, policies and procedures.

d. TIA personnel do not collect data. However, during unit testing TIA personnel must ensure the technology to be tested by the operational user performs in the intended manner. In order to conduct unit testing, TIA personnel will populate the tool to be tested with data that was collected previously by the operational user in accordance with the appropriate governing directives, i.e. EO 12333, DoD 5240.1-R and Army Regulation 381-10. Procedures for handling U.S. person data—like the minimization protocol—are addressed in these governing directives and shall be followed. All persons associated with TIA and TIA-related programs, to specifically include DARPA, contract and contract support personnel with responsibility for TIA

program execution, regardless of location, are prohibited from collecting or otherwise acquiring data.

e. During experiments, DARPA, contract and contract support personnel analyze real data with various tools to examine real problems. The purpose of these experiments is to evaluate the tools for their utility to support the intelligence analysts' mission. As a result of these experiments, interesting results from an intelligence perspective may be generated. Judgments regarding the value of such results and any subsequent production of intelligence is the purview of the operational users and analysts, not DARPA.

f. Any violations of this policy, whether intentional or accidental, shall be reported immediately in accordance with the guidance provided in the implementing instructions.

4. Implementing Instructions:

a. My deputy, Dr. Robert Popp, is responsible for administering the policy concerning the acquisition or use of data about U.S. persons.

b. Implementing instructions in support of this program directive will be issued under separate cover.

A handwritten signature in black ink, appearing to read "John Poindexter". The signature is fluid and cursive, with a long horizontal stroke at the end.

Dr. John M. Poindexter
Director
Information Awareness Office

07 APR 03

Subject: Total Information Awareness (TIA) Program Directive Number 3,
Use of Synthetic Data

From: TIA Program Manager

To: All TIA Contract and Government Personnel Involved with TIA

1. Purpose: The purpose of this TIA Program Directive is to establish an immediate and continuing policy concerning the use of synthetic data.

2. Background: The goal of the Total Information Awareness (TIA) program is to significantly increase the ability of the United States to detect, classify and identify foreign terrorists – and decipher their plans – and thereby enable our nation to take timely action to successfully preempt and defeat terrorist acts. To support this goal the Defense Advanced Research Projects Agency's (DARPA) Information Awareness Office (IAO) developed TIA as a multi-year program consisting of the iterative development, acquisition, testing, refinement and integration of advanced technology and processes. A determination was made to use DoD intelligence entities as test nodes. Other nodes within the Intelligence Community may be established. The US contract support personnel is integral to the TIA program and indeed the program largely consists of contractor personnel executing the program under DARPA guidance and leadership.

3. Program Guidance:

- a. The TIA program will operate within all applicable laws, executive orders and departmental regulations.
- b. All program associated plans and activities conducted by anyone associated with or operating in support of TIA will conform to all applicable laws, executive orders and departmental regulations.
- c. TIA is a research and development activity that for experimentation, development, and demonstration purposes operates within DoD and U.S.G intelligence activities. As such, TIA activities, actions and personnel are subject to all applicable U.S.G and DoD Intelligence Oversight rules, regulations, policies and procedures.
- d. In order to conduct TIA component level testing, synthetic data (i.e., artificial information generated to resemble real-world data) is generated in large data sets. Other simulated data, indicative of terrorist activities, is embedded in these large synthetic data sets. This permits stand-alone testing of these components and avoids using data that might contain information about U.S. persons. However, the use of synthetic data does not preclude the extremely rare possibility that synthetically generated data could resemble a real U.S. person. Effective immediately and in recognition of this possibility, all persons associated with TIA and TIA-related programs, to specifically include DARPA, contract and contract support personnel with responsibility for TIA program execution, regardless of location, are prohibited from distributing synthetic data for any purpose other than the development of TIA or TIA-

related technology. The discovery of synthetic data that resembles real U.S. persons or the distribution of such data for other than TIA development purposes, whether intentional or accidental, shall be reported immediately.

e. Further, all data sets that include synthetic data as well as any work product containing synthetic data shall be marked with the following disclaimer:

Data included herein may be synthetic in nature, i.e. artificial information generated to resemble real-world data. Any resemblance to real persons, living or dead, is purely coincidental. Further dissemination is authorized only as directed by DARPA/IAO or higher DoD authority for use in the development of TIA or TIA-related technology. Release of this information to any other entity or for any other purpose without the consent of DARPA/IAO is strictly prohibited.

4. Implementing Instructions:

a. My deputy, Dr. Robert Popp, is responsible for administering the policy concerning the use of synthetic data.

b. Implementing instructions in support of this program directive will be issued under separate cover.



Dr. John M. Poindexter
Director
Information Awareness Office

07 APR 03

Subject: Total Information Awareness (TIA) Program Directive Number 4,
Memoranda of Agreement with Partners in Experimentation

From: TIA Program Manager

To: All TIA Contract and Government Personnel Involved with TIA

1. Purpose: The purpose of this TIA Program Directive is to establish an immediate and continuing policy concerning the use of Memoranda of Agreement (MOA) with partners in experimentation.

2. Background: The goal of the Total Information Awareness (TIA) program is to significantly increase the ability of the United States to detect, classify and identify foreign terrorists – and decipher their plans – and thereby enable our nation to take timely action to successfully preempt and defeat terrorist acts. To support this goal the Defense Advanced Research Projects Agency's (DARPA) Information Awareness Office (IAO) developed TIA as a multi-year program consisting of the iterative development, acquisition, testing, refinement and integration of advanced technology and processes. A determination was made to use DoD intelligence entities as test nodes. Other nodes within the Intelligence Community may be established. Given potential breadth and depth of TIA capabilities and their usefulness in the greater counter-terrorism and law enforcement communities, it is reasonable to anticipate that partnerships with government agencies will be highly beneficial. These partnering arrangements are considered integral to TIA program objectives to preempt future terrorist attacks.

3. Program Guidance:

g. The TIA program will operate within all applicable laws, executive orders and departmental regulations.

h. All program associated plans and activities conducted by anyone associated with or operating in support of TIA will conform to all applicable laws, executive orders and departmental regulations.

i. TIA is a research and development activity that for experimentation, development, and demonstration purposes operates within DoD and U.S.G intelligence activities. As such, TIA activities, actions and personnel are subject to all applicable U.S.G and DoD Intelligence Oversight rules, regulations, policies and procedures.

d. Effective immediately, any partnership between DARPA and other government agencies for the purposes of advancing TIA research and development through experimentation shall be documented and codified in a formal Memorandum of Agreement (MOA). Each agency partner must ensure that a legal review of the proposed arrangement included in the MOA has been conducted by that agency's legal authority. A coordinating copy of such legal review memo will be forwarded to DoD Office of General Counsel. All parties to any MOA shall execute such agreements prior

to the transfer of any TIA technology or components by DARPA or its supporting TIA and related program contractors.

4. Implementing Instructions:

c. My deputy, Dr. Robert Popp, is responsible for administering the policy concerning Memoranda of Agreement with partners in experimentation.

d. Implementing instructions in support of this program directive will be issued under separate cover.

A handwritten signature in black ink, appearing to read "John M. Poindexter". The signature is fluid and cursive, with a long horizontal stroke at the end.

Dr. John M. Poindexter
Director
Information Awareness Office

07 APR 03

Subject: Total Information Awareness (TIA) Program Directive Number 5,
Resources Provided by Other Government Agencies

From: TIA Program Manager

To: All TIA Contract and Government Personnel Involved with TIA

1. Purpose: The purpose of this TIA Program Directive is to establish an immediate and continuing policy concerning resources provided by other government agencies.

2. Background: The goal of the Total Information Awareness (TIA) program is to significantly increase the ability of the United States to detect, classify and identify foreign terrorists – and decipher their plans – and thereby enable our nation to take timely action to successfully preempt and defeat terrorist acts. To support this goal the Defense Advanced Research Projects Agency's (DARPA) Information Awareness Office (IAO) developed TIA as a multi-year program consisting of the iterative development, acquisition, testing, refinement and integration of advanced technology and processes. A determination was made to use DoD intelligence entities as test nodes. Other nodes within the Intelligence Community may be established. Given potential breadth and depth of TIA capabilities and their usefulness in the greater counter-terrorism and law enforcement communities, it is reasonable to anticipate that partnerships with other government agencies will be highly beneficial. These partnership arrangements are considered integral to TIA program objectives to preempt future terrorist attacks. Such partnerships may also create opportunities for DARPA and TIA contractors to leverage resources from other government agencies.

3. Program Guidance:

- a. The TIA program will operate within all applicable laws, executive orders and departmental regulations.
- b. All program associated plans and activities conducted by anyone associated with or operating in support of TIA will conform to all applicable laws, executive orders and departmental regulations.
- c. TIA is a research and development activity that for experimentation, development, and demonstration purposes operates within DoD and U.S.G intelligence activities. As such, TIA activities, actions and personnel are subject to all applicable U.S.G and DoD Intelligence Oversight rules, regulations, policies and procedures.
- d. The synergies resulting from partnerships with other government agencies may create opportunities for the transfer of resources to DARPA or TIA and TIA-related contractors. Effective immediately and in recognition of the possibility that these resources may contain prohibited data, all persons associated with TIA and TIA-related programs, to specifically include DARPA, contract and contract support personnel with responsibility for TIA program execution, regardless of location, are prohibited from acquiring or using resources from other government agencies known to contain data

that is prohibited by law, executive order, regulation or policy. Any violations of this policy, whether intentional or accidental, shall be reported immediately.

4. Implementing Instructions:

- a. My deputy, Dr. Robert Popp, is responsible for administering the policy concerning the use of resources provided by other government agencies.
- b. Implementing instructions in support of this program directive will be issued under separate cover.

A handwritten signature in black ink, appearing to read "John M. Poindexter". The signature is fluid and cursive, with a long horizontal stroke at the end.

Dr. John M. Poindexter
Director
Information Awareness Office

Appendix E – DARPA–U.S. Army INSCOM Memorandum of Agreement

(Note: This MOA will be used as a model in establishing additional TIA test nodes.)

**MEMORANDUM OF AGREEMENT
BETWEEN THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
AND THE
U.S. ARMY INTELLIGENCE AND SECURITY COMMAND**

PREAMBLE

TECHNOLOGY DEVELOPED BY THE PARTIES TO THIS AGREEMENT SHALL BE USED IN COMPLIANCE WITH AMENDMENT NO. 59 OF PUBLIC LAW 108-2 (WYDEN AMENDMENT). THE PRIMARY PURPOSE OF THIS AGREEMENT IS TO SUPPORT THE LAWFUL ACTIVITIES OF THE DEPARTMENT OF DEFENSE AND NATIONAL SECURITY PROGRAMS CONDUCTED PURSUANT TO LAW.

1. **Purpose.** This Memorandum of Agreement (MOA) establishes the terms of agreement for the Defense Advanced Research Projects Agency (DARPA) and the U.S. Army Intelligence and Security Command (INSCOM) to partner to make advances in the computational and information technologies needed to substantially reduce the threat of international terrorism to the United States.

2. **Background.** The nature of the international terrorist threat does not lend itself to easy identification and exposure by traditional intelligence methodologies. As evidenced by the events of September 11, 2001, it was discovered that international terrorist cells operate in this country and engage in commonplace, everyday events that, when observed in the abstract, do not evidence terrorist activity. Nevertheless, the threat posed by international terrorism to this country cannot now be ignored. Knowledge acquisition and the fusion of singular or very sparse events, within an enormous heterogeneous and dynamic information flow in near real time is key to the detection and comprehension of international terrorist activity. Developing the methodology to extract these events out of a sea of information clutter and integrating them to enable a prediction of future events is the purpose of this MOA.

3. **Responsibilities.** The current risk of international terrorism does not allow a standard research and development (R&D) process for maturing evolving technologies. DARPA will act as INSCOM's R&D partner to accelerate the R&D process and advance new and emerging technologies through the levels of readiness such that sufficiently mature, integrated technologies can be deployed to serve INSCOM and the US intelligence community in the shortest feasible time to facilitate the global war on terrorism. In this regard the parties to this agreement assume the duties and obligations listed below.

a. **DARPA will:**

- (1) **Fiscal.** Comply with fiscal law and policy governing DARPA operations.

**MEMORANDUM OF AGREEMENT BETWEEN THE DEFENSE ADVANCED
PROJECTS AGENCY AND THE U.S. ARMY INTELLIGENCE AND SECURITY
COMMAND**

(2) Personnel. Provide adequate personnel resources to support experimental windows and operational testing.

(3) Equipment. Provide to INSCOM hardware and software resources to adequately support R&D requirements.

(4) Intelligence Oversight. Ensure that all personnel performing functions covered by this Agreement have received the requisite intelligence oversight training. Not participate in any intelligence collection operations or related activities in support of its R&D effort. Serve as the central point of contact to report to the internal and external DARPA oversight boards.

(5) Security. DARPA will serve as the focal point for providing and coordinating the full range of security services necessary to support DARPA program efforts. In accomplishing this task, DARPA security representatives will ensure compliance with INSCOM security policy, practices and procedures.

b. INSCOM will:

(1) Fiscal. Comply with fiscal law and policy governing INSCOM operations.

(2) Space. Provide adequate space within or near the Information Operations Center for DARPA personnel.

(3) Equipment. Provide the basic automation backbone infrastructure required to conduct R&D and application operational activities within the Intelligence Operations Center (IOC) such as access and connectivity to existing networks (NIPRNet, SIPRNet, JWICS, THOR, and HCS), and access to IC Test Net.

(4) Intelligence Oversight. Subject all information provided to DARPA to intelligence oversight and legal review regimes found in Executive Order 12333, DOD Directive 5240.1-R, AR 381-10, and with other US intelligence community policies as required.

(5) Security. Provide all appropriate safeguards for materials and information obtained from DARPA, or others, generated in the course of work under this agreement.

4. Effective Date. This MOA will be effective on the date when the last signatory has signed it and will remain in effect for a period up to five (years) or modified by mutual consent in writing.

5. Modifications. Written modifications to this MOA may be made at any time by INSCOM and DARPA. Modifications shall set forth the exact nature of the change(s)

**MEMORANDUM OF AGREEMENT BETWEEN THE DEFENSE ADVANCED
PROJECTS AGENCY AND THE U.S. ARMY INTELLIGENCE AND SECURITY
COMMAND**

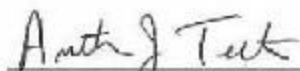
and shall be signed by the original signatories or their successors. No verbal agreement or written statements by anyone other than the signatories or their successors shall be interpreted as modifying or otherwise affecting the terms of this agreement.

6. **Dispute Resolution Procedures.** The parties will try to resolve any disputes under this MOA at the lowest level possible.


7. **MOA Termination.** This MOA will be terminated upon request by either party with at least thirty days notice. All outstanding obligations under the MOA at the time of termination request must be fulfilled before termination.

8. **Review.** This MOA will be reviewed every 120 days by each signatory or their representatives to determine if any changes need to be made.

9. **Approvals.**



ANTHONY J. TETHER, Ph.D.
Director, Defense Advanced
Research Projects Agency



KEITH B. ALEXANDER
Major General, U.S. Army
Commander, U.S. Army Intelligence and
Security Command

April 15, 2003
Date

April 17, 2003
Date