Order Code RL31786

Report for Congress
Received through the CRS Web

Total Information Awareness Programs: Funding, Composition, and Oversight Issues

Updated March 21, 2003

Amy Belasco
Specialist in National Defense
Foreign Affairs, Defense, and Trade Division

Total Information Awareness Programs: Funding, Composition, and Oversight Issues

Summary

Late last year controversy erupted about a Department of Defense (DOD) R&D effort called Total Information Awareness (TIA) under an office headed by retired Admiral John D. Poindexter within the Defense Advanced Research Projects Agency (DARPA). By integrating various new tools designed to detect, anticipate, train for, and provide warnings about potential terrorist attacks, DARPA hopes to develop a prototype Total Information Awareness system. This system would integrate a number of ongoing R&D efforts, referred to in this paper as Total Information Awareness programs. While concern has centered primarily on privacy issues, accounts of the program's funding have also differed. This report covers the funding, composition, oversight, and technical feasibility of TIA programs. The privacy implications are addressed in CRS Report RL31730, *Privacy: Total Information Awareness Programs and Related Information Access, Collection, and Protection Laws,* by Gina Marie Stevens.

In a press interview, Under Secretary of Defense for Acquisition, Technology and Logistics, Edward C. "Pete" Aldridge, stated that the Total Information Awareness project is funded at $10 million in FY2003 and $20 million in FY2004. Other reports indicated higher funding levels of over $100 million in FY2003 and over $200 million for the three-year period, FY2001 - FY2003.

Different accounts of funding levels reflect the fact that DARPA is funding both an integrative effort called the TIA system, as well as 16 individual R&D efforts or TIA programs that could be combined to create that system. In FY2003, DARPA is dedicating $10 million to integrate various R&D efforts into a prototype TIA system, and $137.5 million for the various R&D programs that could make up that system and that are managed by the Information Awareness Office (IAO) headed by Poindexter. Funding for these programs total $137.5 million in FY2003 and $317.0 million for FY2001-FY2003. DOD is requesting $169.2 million for TIA programs in FY2004 and $170.3 in FY2005, and $20 million in FY2004 and $24.5 million in FY2005 for the TIA system integration. These TIA programs are ongoing.

In response to concerns about TIA programs, Congress included special oversight provisions – known as the Wyden amendment – in the FY2003 Consolidated Appropriations Resolution (P.L. 108-7) requiring that the Secretary of Defense, the Director of Central Intelligence and the Attorney General submit a detailed joint report on TIA programs within ninety days or face a cutoff in funding. Senator Feingold, Senator Grassley and other Members also proposed restrictions on data mining in the DOD and the new Department of Homeland Security.

In light of the report required by P.L. 108-7, hearings on TIA programs are likely in the 108th Congress. In addition to privacy concerns, Congress may also address several oversight issues for TIA programs including monitoring collaboration between DARPA and potential users in the law enforcement and intelligence communities and assessing the technical feasibility of the project. This report will be updated as necessary.

Contents

List of Tables

Total Information Awareness Programs: Funding, Composition, and Oversight issues

Current Controversy over Total Information Awareness Programs

Established in January 2002 under retired Admiral John Poindexter, USN, the mission of the Information Awareness Office (IAO) in the Defense Advanced Research Project Agency (DARPA) is to develop new tools to detect, anticipate, train for, and provide warnings about potential terrorist attacks.[1] Within three to five years, DARPA envisions that these tools would be integrated into a prototype Total Information Awareness (TIA) system to provide better intelligence support to senior government officials. If proven effective, Under Secretary of Defense for Acquisition, Technology and Logistics Edward C. "Pete" Aldridge has suggested that the TIA technology prototypes will be turned over to "intelligence, counterintelligence and law enforcement communities as a tool to help them in their battle against domestic terrorism."[2]

In a press conference on November 20, 2002, Under Secretary Aldridge stated that funding for the Total Information Awareness System (TIA) is $10 million in FY2003.[3] On February 7, 2003, he reiterated that funding for the TIA project is $10 million in FY2003 and $20 million in FY2004. The Electronic Privacy Information Center (EPIC), a non-profit organization specializing in privacy issues, calculated that TIA-related programs totaled $112 million in FY2003 and $240 million for the three-year period, FY2001-FY2003.[4] Press reports also cited funding of over $200 million over three years.[5]

These alternative funding levels reflect the difference between the $10 million in funding for the R&D *specifically labeled* the "Total Information Awareness

---

[1] The larger issue of the types of intelligence tools needed to combat terrorism is extensively discussed in Report of the Markle Foundation Task Force, *Protecting America's Freedom in the Information Age*, October 2002; see especially, pp. 25ff, 37ff, 53ff, and 81ff.

[2] Under Secretary of Defense Aldridge as quoted in Defense Department Briefing Transcript, November 20, 2002, p. 10; see [http://www. defenselink.mil].

[3] Under Secretary of Defense Aldridge as quoted in Defense Department Briefing Transcript, November 20, 2002, p. 10; see [http://www. defenselink.mil].

[4] See Electronic Privacy Information Center, "Total Information Awareness (TIA) Budget" on web site: [http://www.epic.org/].

[5] William Safire, *New York Times*, "You are a Suspect," November 14, 2002, see [http://nytimes.com/2002/22/14/opinion/14AF.html]

System" that would integrate various R&D technology efforts, and the $137.5 million in funding for various R&D efforts managed by the Information Awareness Office that could become part of that system. Funding for TIA programs that are managed by the Information Awareness Office includes R&D efforts to develop technologies to improve data mining so as to allow DOD to sift through and analyze patterns in vast amounts of information, to translate large volumes of foreign language materials electronically, to strengthen DOD's information infrastructure, and to devise new tools for high-level decision makers trying to anticipate, train, and respond to terrorist attacks. (See **Appendix** below for descriptions of individual projects).[6]

To proponents, TIA R&D holds out the promise of developing a sophisticated system that would develop new technologies to find patterns from multiple sources of information in order to give decision makers new tools to use to detect, pre-empt and react to potential terrorist attacks. To opponents, TIA has the potential to violate the privacy of individuals by giving the government access to vast amounts of information about individuals as well as possibly mis-identifying individuals as potential terrorists.

Reflecting both these viewpoints, P.L. 108-7 (H.J.Res. 2) the FY2003 Consolidated Appropriations Resolution requires that the Secretary of Defense, the Director of Central Intelligence (DCI), and the Attorney General submit to Congress a detailed report on TIA by May 21, 2003 or face a cutoff in funding (see *Restrictions on TIA in FY2003 Consolidated Appropriations Resolution* later in this report for more details). In the meantime, TIA programs are continuing.[7] DARPA has, for example, obligated $7.4 million of the $10 million available in FY2003 for TIA system integration.[8]

On March 13, 2003, Paul McHale, the new Assistant Secretary of Defense for Homeland Security, testified that although he considered it appropriate for DARPA to develop TIA technologies, once completed, DOD did not anticipate using the technology because of the desire that " this kind of intrusive but perhaps essential capability" be operated by civilian rather than military personnel.[9] Instead, he anticipated that the TIA system would be transferred to civilian law enforcement agencies and be subject to the judicial and congressional oversight.[10]

---

[6] See description of TIA in DARPA, RDT&E Descriptive Summaries for FY2003 (or the R-2), available at the DARPA web site: [http://www.dtic.mil/comptroller/fy2003budget/budget_justification/pdfs/rdtande/darpa_vol1.pdf]

[7] Press reports indicating that TIA programs have been terminated are inaccurate.

[8] Information provided to CRS by DARPA, February 2003.

[9] Testimony of Paul McHale before the Subcommittee on Special Oversight Panel on Terrorism, Unconventional Threats and Capabilities, House Armed Services Committee, Hearing on Force Protection, March 13, 2003.

[10] *Ibid*.

FY2001-FY2003 Funding Levels

According to DARPA, technology developed in some or all of the sixteen R&D efforts managed by the Information Awareness Office may be integrated into the TIA system.[11] DARPA's FY2003 request for the R&D efforts managed by the Information Awareness Office totaled $137.5 million in FY2003 (see **Table 1** below), including $10 million for the integrative efforts *specifically labeled* the Total Information Awareness System, a new start in FY2003.

Technology Currently Linked to the TIA System. DARPA's FY2003 budget materials state that TIA will integrate technology and components from at least 8 of the 16 R&D efforts (including the integration itself) that are managed by the Information Awareness Office.[12] According to DARPA, TIA is "the assured transition of a system-level prototype that integrates technology and components developed in other DARPA programs *including* [italics added] Genoa and Genoa II ... TIDES ..., Genisys, EELD, WAE, HID, and Bio-Surveillance ... "[13] (See **Table 2** and the **Appendix** for funding and description of these R&D efforts).

Funding for these eight R&D efforts totals $110.6 million in FY2003, $83.8 million in FY2002, and $65.0 million in FY2001 (see **Table 1**). Three follow-on machine translation efforts under the Information Awareness Office will probably also be incorporated into the TIA system.

Information Awareness Office-Managed R&D. According to DARPA, the TIA system may also exploit the results of other R&D efforts that are under the Information Awareness office, other DARPA efforts, or R&D conducted outside of DARPA.[14] Several DARPA R&D efforts under other offices appear to have similar purposes to those specifically linked to TIA.[15] DARPA also hopes to exploit commercial data mining technology and R&D developed by other agencies like the

---

[11] See table and appendix for how R&D linked to TIA is shown in DARPA's budget justification materials. DARPA provided CRS with the list of 16 R&D efforts that are managed by the Information Awareness Office.

[12] Eight counts Genoa and Genoa II as one project, and includes TIA integration as one of the components.

[13] See description of TIA in DARPA, RDT&E Descriptive Summaries for FY2003; see [http://www.dtic.mil/comptroller/fy2003budget/budget_justification/pdfs/rdtande/darpa_vol1.pdf]

[14] See Briefing by John Poindexter, Director, Information Awareness Office, to Congressional Authorizing Committees Staff, February 26, 2002. For example, DARPA spokesman suggested that TIDES system could be combined with OASIS, a system designed to protect DOD's information systems from cyber attack; see 23rd DARPA System and Technology Symposium July 29-August 2, 2002 on web site shown below. [http://www.darpa.mil/DARPATech2002/presentation.html].

[15] For example, other DARPA offices manage Software for Situational Analysis and Rapid Knowledge Foundation, two programs designed to find ways to exploit multiple data bases, in this case to identify biowarfare threats, just as Genisys and EELD, two TIA-linked efforts, analyze and mine data to identify potential terrorists.

National Security Agency. According to the Director of DARPA, all funding managed by the Information Awareness Office is considered to be Total Information Awareness programs.[16]

Funding for projects managed by the Information Awareness Office totals $137.5 million in FY2003, $99.5 million in FY2002, and $80 million in FY2001. Over the three-year period, FY2001- FY2003, funding totals $317.0 million. The increase in FY2003 reflects several new starts in FY2003 for Genisys, a comprehensive data mining effort, MIDGET, a system designed to prevent contamination of open databases, Rapid Analytic Wargaming, a tool for decision makers, and the TIA integration effort (see **Table 2** below and **Appendix**).

Table 1. Funding for Information Awareness Office and for Total Information Awareness Technology, FY2001-FY2003

(in millions of dollars)

Total Funding	**FY 2001**	**FY 2002**	**FY 2003[a]**	**FY 01-03**
Total Information Awareness System integration[b]	0.0	0.0	10.0	10.0
Technology Programs Currently Linked to TIA [c]	65.0	83.8	110.6	259.4
Information Awareness Office programs	80.0	99.5	137.5	317.0

Sources and Notes:

See DARPA, RDT&E Descriptive Summaries for FY2003 (or the R-2), available at web site, [http://www.dtic.mil/comptroller/fy2003budget/budget_justification/pdfs/rdtande/darpa_vol1.pdf]

[a] FY2003 level reflects DARPA's request.

[b] TIA is shown by DARPA as a specific R&D effort in Project CCC-01 in Program Element 603760E.

[c] Includes the 8 R&D efforts identified in DARPA's FY2003 budget justification materials as specifically linked to the TIA system, including four data mining efforts (Human Identification at a Distance, Evident Extraction and Link Discovery, Genisys, Bio-surveillance), machine translation of languages (TIDES), and three decision making tools (Wargaming the Asymmetric Environment, Project Genoa/Genoa II, and Total Information Awareness); see appendix for description of these efforts.

Although the TIA system was first proposed as an integrated entity in the FY2003 budget shortly after establishment of the Information Awareness Office, some of the R&D efforts that could become part of that system have been underway for a number of years. In fact, several of the R& D efforts, e.g. Project Genoa and machine translation of languages, first received funds in 1996 and 1997 respectively. For comparative purposes, **Table 1** above and the more detailed **Table 2** below show funding from FY2001 through FY2003 for all the elements now managed by the Information Awareness Office that could become part of the Total Information Awareness system.

---

[16] Statement during briefing to congressional staff by Dr. Tony Tether, Director of DARPA, "DARPA's Information Technology Initiative on Countering Terrorism, January 27, 2003. on January 27, 2003.

Authorization and Appropriation of DOD RDT&E Programs. Funding for DARPA, as for the Research, Development, Test & Evaluation (RDT&E) programs of the services, is authorized and appropriated annually at the account level. In the case of DARPA, funding is included within the RDT&E, Defensewide account.[17] The TIA system, like other R&D efforts, is not specifically identified in statutory language in the FY2003 DOD authorization or appropriation acts.

Congressional intent about the funding levels for individual R&D efforts, however, may be included in committee reports, and is considered binding. The FY2003 DOD authorization and appropriation conference reports did not include any specific language about the TIA system, and the House and Senate appropriators voiced different views about various Total Information Awareness components.[18]

FY2001-FY2003 Funding for Individual R&D Efforts. Based on their primary purpose, the sixteen R&D efforts managed by the Information Awareness Office have been grouped into the four categories below. **Table 2** below shows the funding for FY2001-FY2003 for the individual R&D efforts managed by the Information Awareness Office, including those R&D efforts currently designated as part of the TIA system.[19] The **Appendix** briefly describes each R&D efforts.

New Data Mining and Analysis Technologies. These R&D efforts are designed to develop technologies that would be capable of sifting through large data bases, e.g. financial, communications, travel, to detect patterns associated with terrorists' activities. Total funding for these efforts was $29.2 million in FY2001, $38.2 million in FY2003 and $53.0 million in FY2003. Increases reflect initiation

---

[17] DARPA provides detailed descriptions of its programs and projects in budget justification materials submitted to Congress annually.

[18] The FY2003 appropriation conference report mentions only one TIA component, Genisys, suggesting that delays might justify lower funding; see Committee of Conference on Appropriations, *Making Appropriations for the Department of Defense for the Fiscal Year Ending September 30, 2003, and for other purposes*, H.Rept. 107-732, p. 305. The House and Senate versions of the FY2003 DOD Authorization Act made different recommendations about Program Element 0602301E, which funds some of the R&D managed by IAO. The House recommended no reductions and commended DARPA's overall information awareness programs, and the Senate recommended cuts in two R&D efforts under IAO, the Bio-Surveillance and Genisys R&D efforts. For House action, see House Armed Services Committee, *Bob Stump National Defense Authorization Act for Fiscal Year 2003*, May 3, 2002, H.Rept. 107-436, p. 239 and p. 241. For Senate action, see Senate Armed Services Committee, *National Defense Authorization Act for Fiscal Year 2003*, May 14, 2002, S.Rept. 107-151, p. 230.

[19] The 16 R&D efforts have been grouped into categories based on Department of Defense, *FY2003 Budget Estimate, Research, Development, Test and Evaluation, Defense-wide, Volume 1, Defense Advanced Research projects Agency*, and briefings by project managers to the 23rd DARPA System and Technology Symposium, July 29 - August 2, 2002; see [http://www.dtic.mil/comptroller/fy2003budget/budget_justification/pdfs/rdtande/darpa_vol1.pdf] and [http://www.darpa.mil/DARPATech2002/presentation.html]. Table 2 in this report shows how the various TIA components are included in program elements and projects in DARPA's FY2003 Budget Estimate.

of the Bio-surveillance effort in FY2002 and the Genisys program in FY2003, both of which have raised privacy concerns.

New Machine Translation Technologies. These R&D efforts are intended to develop new software technology to translate large volumes of foreign language material, both written and oral, that would be collected from sources ranging from electronic sources to battlefield transmissions. At $36 million annually, funding for these efforts was stable between FY2001 and FY2003.

Protection of Critical Information Infrastructure. These R&D efforts are intended to protect DOD's information infrastructure and detect mis-information in open-source data that DOD may collect. Funding in this area grew from zero in FY2001 to $2.0 million in FY2002 with the initiation of DefenseNet, and jumped to $9.5 million with the new Mis-Information Detection and Generation effort.

Tools for High-Level Decision Makers. These R&D efforts are intended to develop tools, ranging from war-gaming simulations to collaborative reasoning processes, designed to help high-level decision makers anticipate, train for, pre-empt, or react to terrorist acts. Funding for these efforts increased from $14.4 million to $23.5 million in FY2002 with the doubling in the funding level for Wargaming the Asymmetric Environment. That funding jumped to $39.5 million with the initiation of Total Information Awareness System, the integrative effort.

Table 2. FY2001-FY2003 Funding for Information Awareness Office and Total Information Awareness Programs

(In millions)

Major Purpose by Category	Project	Program Element	FY01	FY02	FY03 Request
Data Mining and Analysis Technologies Subtotal:			29.2	38.2	53.0
Human ID at a Distance*	Asymmetric Threat ST-28	602301E	11.8	15.9	14.5
Evidence Extraction and Link Discovery (EELD)*	ST-28	602301E	17.3	14.4	14.0
Genisys*	ST-28	602301E	0.0	0.0	11.0
Bio-surveillance*	ST-28	602301E	0.0	8.0	13.5
Machine Translation of Languages Subtotal:			36.5	35.8	35.5
Translingual Information Detection, Extraction and Summarization (TIDES)* and Effective, Affordable, Reusable Speech-to-Text (EARS), and Multispeaker Environments (MUSE) and Global Autonomous Language Exploitation (GALES)[a]	Intelligent Systems and Software, ST-11	602301E	21.5	22.1	22.1
Babylon and Communicator[a]	ST-11	602301E	15.0	13.7	13.4
Protection of DOD's Information Infrastructure Subtotal:			0.0	2.0	9.5
DefenseNet (DNET)[b]	ST-28/ST-11[b]	602301E	0.0	2.0	3.0
Mis-Information Detection and Generation (MIDGET)	ST-28	602301E	0.0	0.0	6.5
Tools for High-Level Decision makers Subtotal:			14.4	23.5	39.5
Rapid Analytic Wargaming (RAW)	ST-11	602301E	0.0	0.0	4.0
Wargaming the Asymmetric Environment (WAE)*	Command & Control Info. Systems, CCC-01	603760E	6.9	15.8	18.5
Project Genoa/Genoa II [c]*	CCC-01	603760E	7.5	7.6	7.0
Total Information Awareness System [d]*	CCC-01	603760E	0.0	0.0	10.0
Technology Supporting TIA System*	NA	NA	65.0	83.8	110.6
Three-Year Total, FY2001-FY2003*	259.4				
Information Awareness Office Total	NA	NA	80.0	99.5	137.5
Three-Year Total, FY2001-FY2003:	317.0				

Sources and Notes :

DARPA and Total Information Awareness Office program: [http://www.defenselink.darpa.mil/iao/programs].
See DARPA, RDT&E Descriptive Summaries for FY2003 (or the R-2), available at web site, [http://www.dtic.mil/comptroller/fy2003budget/budget_justification/pdfs/rdtande/darpa_vol1.pdf].

* identifies R&D linked specifically by DARPA to TIA System

[a] Funding for individual components not shown in DARPA's FY2003 budget justification.

[b] DefenseNet transfers from Project ST-28 in FY2002 to Project ST-11 in 2003; see DARPA's R-2, p. 90; or, [http://www.dtic.mil/comptroller/fy2003budget/budget_justification/pdfs/rdtande/darpa_vol1.pdf].

[c] Funding for GenoaII starts in FY2003.

[d] Total Information Awareness is the integrative effort.

Future Funding for Information Awareness Office Programs

For FY2004, DARPA is requesting $169.2 million for TIA programs and $170.3 million in FY2005.[20] If DARPA funds the R&D efforts that are managed by the Information Awareness Office comparably to funding in previous years, annual funding for TIA programs would average about $145 million annually.[21] The higher levels requested by DOD in the FY2004 budget suggest additional emphases by DARPA on this program. If past funding trends hold, DARPA could spend about $600 million for TIA-related R&D in the next four years, at which point the project is slated to be complete. This funding would be in addition to the $317 million spent from FY2001-FY2003.

Ongoing DARPA Collaboration

DARPA's goals for TIA programs call for sharing of information and analysis among DOD, the intelligence community, counter-intelligence, law enforcement and high-level policy and operational decision makers who could exploit both commercial data mining and analysis systems and new tools being developed in TIA programs. DARPA has also consulted with other DOD offices, such as Strategic Command.[22] Thus far, DARPA's collaboration with agencies outside DOD has been informal, including an unsigned memorandum of understanding developed with the FBI and meetings with Office of Homeland Security officials.[23]

Within DOD, DARPA has established a site at the Army's Information Dominance Center at Fort Belvoir to test potential elements of the TIA system, such as Genoa, by applying various tools in an operational environment using data about U.S. persons that is available to the intelligence community under existing laws and policies. That information includes 13 categories of information ranging from publicly available data to information about potential intelligence sources.[24]

---

[20] DARPA, "Paper in response to questions from CRS," February 2003. DOD submits a two-year budget but Congress appropriates only one year of funding

[21] DARPA's FY2003 budget justification material includes funding estimates for FY2004 - FY2007 at the project level. The average share of TIA-related R&D in the relevant projects for FY2001-FY2003 can be used to project funding levels for future years. For example, all funding in DARPA for Project ST-28, Asymmetric Threat in Program Element 0602301E, which is solely dedicated to TIA-related projects, can be included. In addition, about half of the funding in Project ST-11, Intelligent Systems and Software, and about 15% of the total for Project CCC-01, Command & Control Information Systems in PE 0603760E, may also be dedicated to TIA based on their shares in earlier years.

[22] Briefing by Dr. Tony Tether, Director, DARPA to Congressional staffers, "DARPA's Information Technology Initiative on Countering Terrorism, January 27, 2003.

[23] DARPA's Director, Dr. Tony Tether, stated that DARPA has a draft unsigned MOU with the FBI during the January 27, 2003 briefing to Congressional staffers.

[24] DOD Regulation 5241.1-R, Procedure 2 lists 13 types of information about U.S. persons that DOD intelligence components are permitted to collect: information obtained with consent, that is publicly available, foreign intelligence, counterintelligence, sources that

(continued...)

DARPA is also testing other potential TIA components, like Genisys, by using fictitious data and mock "Red" or terrorist teams who create potential terrorist scenarios, as well as experimenting with linking its intelligence information with a variety of commercially available data mining systems as and systems developed by other government agencies like the National Security Agency.[25] Through these various experiments, DARPA hopes to test the utility of various data mining tools in identifying potential terrorists. In addition, DARPA has tried out some of its tools on information obtained from prisoners at the U.S. naval base at Guantanamo, Cuba.

Restrictions on TIA in FY2003 Consolidated Appropriations Resolution and Other Legislative Proposals

The FY2003 Consolidated Appropriations Resolution, P.L. 108-7 (H.J.Res. 2) includes a provision requiring that the Secretary of Defense, the Attorney General and the Director of Central Intelligence submit a joint, detailed report to Congress within ninety days or face a cutoff of funding. These restrictions on TIA were originally proposed by Senator Wyden. The required report on TIA programs is to:

- explain and show planned spending and schedules for each TIA project and activity;
- identify target dates for deployment of each component;
- evaluate the system's likely effectiveness in predicting terrorist activities;
- assess the likely impact of implementation on privacy and civil liberties;
- list laws and regulations governing collection efforts and identify any changes that would be needed with deployment of TIA; and
- include recommendations from the Attorney General about procedures, regulations or legislation that would eliminate or minimize adverse effects of any TIA programs on privacy and civil liberties.[26]

If no report is submitted, the funding cutoff can be avoided if the President certifies in writing to Congress that submitting the report is not practicable and that ending R& D on Total Information Awareness programs would endanger national security.

---

[24] (...continued)
could assist intelligence, sources that could help identify or protect intelligence information, information about potential suspects threatening DOD security, personnel security investigations, communications security investigations, narcotics suspects, threats to safety, information available from general overhead reconnaissance, and collected for administrative purposes. See following web site for this and related regulations: [http://www.dtic.mil/whs/directives/corres/pdf/d52401_042588/d52401p.pdf].

[25] DOD Briefing Transcript, November 20, 2002; [http://www.defenselink.mil].

[26] See Division M, Section 111 of H.J.Res. 2 in *Congressional Record*, February 12, 2003, Part Two.

In addition, the provision requires that DOD notify Congress and receive specific appropriations and authorization for any deployment or transfer to another federal agency of any TIA component unless the component is to be used for overseas military operations or for foreign intelligence activities conducted against non-U.S. persons.[27]

Other Members of Congress have also signaled concerns about the TIA system. On January 16, 2003, Senator Feingold and others introduced S. 188, the Data Mining Moratorium Act of 2003 that would place restrictions on data mining activities in DOD and other agencies. In November 2002, Senator Grassley asked the DOD Inspector General to conduct an audit of TIA programs and asked Attorney General Ashcroft to provide by February 10, 2003 information about any involvement that the Department of Justice or the FBI have had with the TIA program. Senator Grassley has not yet received a reply.[28]

Issues for Congress

In addition to concerns raised by members of Congress and public interest groups about protecting the privacy of U.S. citizens, Congress may continue to address oversight issues, including:

- developing monitoring mechanisms for TIA programs; and
- assessing the technical feasibility of the program.

Monitoring TIA Programs

DARPA suggests that its role in developing prototype technologies for a TIA system is consistent with both its mission and history of sponsoring basic research for the mid and long-term that crosses service lines, and has multiple potential users, both inside and outside DOD. Previous examples of DARPA-developed technology with wide-ranging implications include stealth technology, Global Positioning System (GPS), and development of the Internet.[29] Based on recent testimony by Assistant Secretary of Defense Paul McHale emphasizing that DOD did not expect to use a TIA system but would turn the system over to civilian law enforcement

---

[27] The final version changes the original Wyden amendment (SA59) by extending the amount of time for submission of the report from sixty to ninety days and by clarifying that TIA components could be used in the U.S. if they were applied to non-U.S. persons. See *Congressional Record*, January 17, 2003, p. S1165 for original version of the Wyden amendment; compare to H.Rept. 108-10 on H.J.Res. 2, FY2003 Consolidated Appropriations Resolution in *Congressional Record*, February 12, 2003, Book Two. For the changes to the Wyden amendment, compare Division M, Section 111 (a) (1) and (c) (2) (B).

[28] Senator Chuck Grassley, *Press Release*, January 21, 2003, and conversation with Judiciary Committee staff, March 12, 2003.

[29] Tether briefing, January 2003.

agencies, TIA may not have a defense mission.[30] In describing plans for the TIA system, DARPA's Director, Dr. Tony Tether, cited collaboration with potential users in other federal agencies as a key part of their approach.[31]

Yet that collaboration – between the law enforcement community and the intelligence community, for example – has raised concerns among some observers about the roles of different agencies in gathering and sharing intelligence on potential threats from terrorists located in the United States. Those concerns reflect the experiences of the 1960s and 1970s when the FBI's counterintelligence program targeted civil rights and anti-war organizations as part of its efforts to pursue domestic terrorists.[32]

DARPA's efforts at collaboration reflect the fact that there are potentially many users of any tools that DARPA develops to predict terrorist threats. Currently, several agencies are or will be collecting or analyzing intelligence on potential terrorist threats, including the Counterterrorist Center under the CIA, the FBI's Joint Terrorist Task Forces, the new Department of Homeland Security. Another new user would be President Bush's proposed new Terrorist Threat Integration Center to be established May 1, 2003 with the mission of integrating all of U.S. government information and analysis about potential terrorist threats.[33] DARPA envisions working with potential users in the design of its tools for decision makers, a practice, that could be difficult with restrictions on transfer of TIA components.

Sharing information among several users makes it more difficult to protect both intelligence sources and the privacy of individuals. For that reason, DARPA is sponsoring some research on developing 'fire walls' that would protect the sources of intelligence gatherers and prevent potential leakage among users. The distributed type of system that DARPA envisions could make those challenges greater. Early collaboration with potential users, for which DARPA has been praised, could also create problems with ensuring privacy and preventing misuse of intelligence sources and data on individuals, particularly if DARPA tries to exploit multiple data bases and to share data across agencies.[34]

Developing tools to ensure that the privacy of both sources and individuals is both a technical challenge and a policy issue. DARPA's Genisys program, a TIA component intended to integrate and query large data bases that has raised privacy concerns, also includes R&D on tools to ensure privacy. These tools may include

---

[30] Testimony of Paul McHale before the Subcommittee on Special Oversight Panel on Terrorism, Unconventional Threats and Capabilities, House Armed Services Committee, Hearing on Force Protection, March 13, 2003.

[31] Briefing by Dr. Tony Tether, Director, DARPA to Congressional staffers, "DARPA's Information Technology Initiative on Countering Terrorism, January 27, 2003.

[32] Markle Foundation report, " A Primer on the Changing Role of Law Enforcement and intelligence in the War on terrorism," by Robert M. McNamara, Jr., p. 85.

[33] See CRS Report RS21283, *Homeland Security: Intelligence Support* by Richard Best.

[34] Report of the Markle Foundation Task Force, *Protecting America's Freedom in the Information Age*, October 2002, p. 14-15, 22, 26, and 27.

"partitioning," which segregates transactions from the identity of the individual, filters to limit access to information and software agents that would delete unrelated information. According to a technical group tasked by DARPA to look into technological solutions to privacy issues, the Information Science and Technology panel (ISAT), there are significant difficulties in developing tools and protocols to protect privacy. This group called on DARPA to devote significant research resources in this area, and to establish a citizen advisory board to privacy policy standards.[35]

On February 7, 2003, the Department of Defense established two boards to monitor TIA programs.[36] Made up of high-level DOD officials, the internal TIA oversight board is tasked with setting policies and procedures for use of TIA tools within DOD and establishing protocols for transferring TIA capabilities outside of DOD to ensure consistency with privacy laws and policies. DOD also established an outside advisory board including experts in privacy issues, to advise the Secretary of Defense on policy and legal issues raised by using advanced technology to identify and predict terrorists threats.[37] In separate statements to reporters, Senator Wyden and a spokesman for the American Civil Liberties Union each suggested that the new boards proposed by the Pentagon did not eliminate the need for Congressional oversight."[38]

P.L. 108-7, passed by both houses the following week, requires that DOD inform and get Congressional authorization for any transfers between agencies or for deployment of any TIA components. Under P.L. 108-7, testing outside of DOD may also be subject to rigorous oversight. In its current research, DARPA has been careful to use 'dummy' or fictitious data on individuals to test the effectiveness of various models for detecting potential terrorists, or to use only data that is currently legally permissible for intelligence gathering purposes (see discussion of ongoing DARPA collaboration above). If DARPA's technology efforts - in data mining or model development - are to be fully tested, however, real data, with all its flaws, may need to be used, and using real data may raise privacy issues. To decrease the potential for significant errors in the prototype models and systems under development, extensive testing efforts could be desirable.

---

[35] Information Science And Technology (ISAT) study Group, *Security with Privacy*, 13 December 2002.

[36] See, DOD Press Release, " Total Information Awareness Update, February 7, 2003; see [http//www.defenselink.mil/news/Feb2003/b02072003_bt060-03.html]

[37] DOD Press Release, "Total Information Awareness (TIA) Update," February 7, 2003. Members of the advisory board would be Newton Minow, Northwestern University, Zowe Baird, president Markle Foundation, Floyd Abrams, civil rights attorney, Gerhard Casper, Former president of Stanford University, Griffin Bell, former U.S. Attorney General and judge, William T. Coleman, CEO of BEA, Lloyd Cutler, former White House Counsel.

[38] *New York Times*, "Pentagon Forms 2 Panels To Allay Fears on Spying," February 8, 2003; *Boston Globe*, "2 Panels to Monitor Eavesdropping, Pentagon Hopes to Assuage Critics of Defense Plan," February 8, 2003.

Assessing Technical Feasibility

While some observers see great potential in DARPA's TIA proposals to exploit a wide range of data bases and develop models to identify terrorists, other observers are skeptical even models with sophisticated algorithms could pick terrorists out from large data bases, the proverbial problem of finding a needle in a haystack. DARPA's description suggests that the TIA system will be developed using a variety of data mining techniques coupled with models developed by analysts. Although there does not appear to be any simple definition, data mining has been defined as exploiting a variety of tools to extract predictive information from large data bases.[39]

Several major technical problems are inherent in data mining and model development that would need to be solved to develop an effective TIA system including:

- identifying and getting access to appropriate data bases;
- cleaning up "dirty" or inaccurate data in data bases;
- integrating disparate data bases;
- developing models or algorithms to identify likely terrorists;
- mis-identifying suspects because of large numbers of false leads; and
- dealing with timing and cost dilemmas.

Data Base Problems. Getting access, 'cleaning up,' and integrating large data bases may pose significant challenges in developing a TIA system. While DARPA is currently looking at links between military intelligence data and other sources at its Army testing site, there could be complications in linking to other data bases and ensuring that only permissible data is included.[40] In addition, any data base includes a significant number of errors – a problem routinely discussed by data mining experts – and it is not clear that there are adequate methods for catching errors. Linking large and disparate data bases is not only a challenging task in itself but could compound the number of errors.

Searching large data bases with large numbers of errors could both reduce the likelihood that terrorists would be identified and magnify the possibility that individuals who are not terrorists would be tagged. Erroneous data may be included either inadvertently by those entering the data or intentionally by "identity threat" where individuals deliberately impersonate others, worrisome problems to technical and privacy experts alike. The quality of the data could be diluted further if disparate data bases are linked.

Developing Ways To Identify Terrorists. DARPA plans to use both quantitative and qualitative data mining techniques to develop tools to identify terrorists. Data mining techniques are currently widely used for commercial purposes, ranging from targeted marketing to detecting credit card fraud, as well as

---

[39] See Puhpa Ramachandran M, *Mining for Gold*, White Paper, December 2001.

[40] Letter from Barbara Simons, Ph.D., and Eugene H. Spafford, Ph.D, Co-Chairs, U.S. Association for Computing Machinery to Senators John Warner and Carl Levin, Senate Armed Services Committee, January 23, 2003. See [http://www.acm.org/usacm].

for law-enforcement (e.g., to catch drug smugglers). In these cases, however, analysts and statisticians develop, test and re-test algorithms or quantitative relationships in order to hone formulas and improve their accuracy in detecting patterns. In the case of credit card fraud, for example, statistical algorithms or pattern identifying techniques can be refined with follow-up checks of billing records.

According to DARPA's descriptions, TIA components would develop technologies using both statistically-based algorithms to detect patterns in multiple data sources from a wide range of sources – financial, telephonic, foreign messages, intelligence traffic – and models of terrorist behavior based on analysis of historical experiences and scenarios developed by analysts. DARPA anticipates that by speculating, analysts will develop scenarios of particular terrorist attacks and then back into the types of activities that would be necessary to carry out those attacks. Some observers have suggested that it could be difficult to anticipate terrorist acts, and our success in anticipating previous terrorist attacks has been limited. With the enormous increases in the speed of processing information and the proliferation of data mining techniques, DARPA sees new opportunities for exploiting a variety of information sources using quantitative techniques like data mining.

Technology experts and others, however, have questioned whether the problem of detecting potential terrorists is susceptible to the data mining techniques routinely done by commercial companies in light of the difficulty in predicting terrorist behavior. The problem is made all the more difficult by the likelihood that the number of Al Qaeda members in the U.S. is small; a widely-quoted FBI estimate of 5,000 was later dismissed as too high, a small number compared to the large number of transactions that are analyzed in commercial data mining applications.[41]

In response, DARPA suggests that its research would not simply search data bases for potential terrorists but instead would develop templates, based on studies of past attacks and captured terrorists documents, that would be used to focus searches of databases more narrowly. In addition, the process would be iterative, in other words, analysts would use a variety of techniques, sequentially, to identify potential terrorists.[42]

The Problem of False Leads. A key element in assessing the viability of the TIA system is whether the technologies developed will be sufficiently accurate to limit the number of potential suspects and minimize the number of false leads so as to avoid misidentifying individuals as suspects.[43] If the number of potential suspects or false leads proves to be large, the timeliness of warnings, as well as the cost of conducting followup checks, could also make a TIA system problematic. Some observers are also concerned that if DOD or intelligence agencies identified significant numbers of false leads, the pressures of time and urgency could lead to violations of the rights of individuals.

---

[41] New York Times, "5,000 Al Qaeda Operatives in The U.S.," February 16, 2003.

[42] Briefing to Congressional Staff by Dr. Tony Tether, *DARPA*, January 2003.

[43] Shane Harris, *Government Executive*, "Total Information Awareness official responds to criticism," January 31, 2003.

DARPA contends that concerns about false leads (called false alarms or "false positives" by statisticians) are exaggerated. In credit card fraud, for example, a false alarm or false positive would mistakenly identify a transaction as fraudulent. To avoid false alarms, DARPA argues that a TIA system would use multiple means to identify suspects, ranging from models developed by "Red Teams" envisioning terrorist scenarios to patterns detected by linking intelligence data with commercially developed data mining techniques. Using such a tiered approach, DARPA contends that suspects would only be tagged after multiple checks.

Some observers have questioned whether these techniques could successfully cull the number of suspects. But assuming that DARPA's approach could reduce the number, capturing a certain number of false leads is inherent in statistical techniques. For example, consider the extensive work of the credit card industry in developing techniques to identify credit card fraud. In a controlled trial, researchers tested the effectiveness of combining several statistical tools to identify credit card fraud using a large, real testing sample of 500,000 transactions, deliberately seeded with 100,000 fraudulent transactions in order to refine statistical algorithms.[44] (See **Table 3**).

Table 3. Illustrative Credit Card and Terrorist Cases

Sample Credit Card Data Base	Illustrative Terrorist Data Base
Total data base: 500,000 transactions, including:	Total data base: 1,000,000 transactions, including:
Fraudulent: 100,000[a]	Terrorists: 5,000[a]
Legitimate: 400,000	Other suspects: 995,000
Tagged as fraudulent: Sum of	Tagged as terrorists: Sum of
Actual fraudulent: 50,000 (100,000 x .5):	Actual terrorists: 1,500 (5,000 x .3)
False alarms: 80,000 (400,000 x .2)	False Alarms : 298,500 (995,000 x .3)
Total cases to be investigated: 130,000	Total cases to be investigated: 300,000
Ratio of suspects to fraudulent: 2.6:1	Ratio of suspects to terrorists: 200:1

Sources: CRS example developed based on discussions with member of Association of Computing Machinery, and Stolfo, Fan, Prodromidia, and Chan, "Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results;" for paper, see following web site: [http://www.cs.fit.edu/~pkc/papers/].

Note: [a] Examples assume an incidence rate for wrongdoers of 20% for the credit card example and 1/2% for the terrorist data base.

---

[44] Researchers have to know the composition of the data in order to test the effectiveness of their tools. These examples were developed by CRS with the help of a member of the Association for Computing Machinery using the article, Stolfo, Fan, Prodromidia, and Chan, "Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results; " see paper on following web site: [http://www.cs.fit.edu/~pkc/papers/].

The researchers found that by combining several statistical tools, they could catch about 50% of the actual fraudulent transactions with a false alarm rate of about 20%. In other words, while 50,000 of the fraudulent cases were identified, (50% of 100,000), another 80,000 cases were mistakenly tagged as fraudulent (20% of 400,000 legitimate transactions) at the same time. Investigators therefore would need to investigate 130,000 cases to catch 50,000 wrongdoers, or about 2.6 cases for every 1 wrongdoer. In the case of credit card fraud, algorithms have been extensively refined using large amounts of real data, and followup checks on leads are routine as anyone who has received a phone call after making an unusually large charge knows.

Even in the case of credit card fraud, however, the incidence of wrongdoers is likely to be below 20%. (The actual fraud rate is a closely-guarded industry secret.) When the incidence of fraud is lower, the chances of identifying wrongdoers decreases.[45] Press reports last summer cited an FBI estimate of 5,000 Al Qaeda operatives in the U.S., but that estimate was later dismissed by the government, and experts suggested that hundreds rather than thousands was the more likely number.[46] In light of the relatively small number of terrorists, the likelihood of catching them, even with targeted data bases, could be far lower. The chance, as well as the cost to individuals of mis-identifying suspects, could also be far greater.

An illustrative case using statistical algorithms to identify terrorists that would increase the chances that a TIA system would work could be based on the following assumptions:

- the data base would be limited to 1,000,000 transactions because DARPA had successfully culled the number of suspects; and
- there are 5,000 terrorists in the data base, an incidence rate of 1/2 %.

The number of terrorists to be identified would then be 5,000 (1/2% of 1,000,000).

At the same time, assume optimistically that a combination of data mining and modeling tools could identify 30% or 1,500 of the 5,000 terrorists but that the false alarm rate was 30% because the difficulty of identifying terrorists is greater than detecting credit card fraud. In this case, investigators would need to check a total of 300,000 cases to catch the 1,500 terrorists (30% of 5,000 terrorists + 30% of 995,000 other suspects). For every terrorist identified, some 200 other suspects would have to be investigated.

Some computer experts think that even this case is optimistic. If DARPA's data base was larger, the number of false alarms could be far greater, even with a high accuracy rate. In examples proposed by computer experts that assumed a highly accurate TIA system was applied to the entire U.S. population, the number of false

---

[45] *Ibid*. In this research case, the fraud catching rate drops from 80% to 50% when the incidence of fraud decreases from 50% to 20%.

[46] *New York Times*, "5,000 Al Qaeda Operatives in The U.S," February 15, 2003, and *Washington Times*, "5,000 in U.S. Suspected of Ties to Al Qaeda," July 11, 2002.

alarms could be 3 million people annually.[47] Either case would pose considerable challenges to investigators, particularly in cases where a threat was considered imminent. If the number of potential suspects identified was significant, the cost of implementing the system could also grow, as substantial personnel would be needed to investigate potential leads and ensure that false leads were eliminated.

Appendix: Description of R&D Efforts Managed by the Information Awareness Office By Category

(* = R&D efforts specifically linked to the TIA system by DARPA)

Data Mining Technologies.

- **Human Identification at a Distance (HumanID).*** This project aims to use information from sensors about human characteristics such as gait or face, to identify individuals at any time of the day or night and in all weather conditions, for instance, within a large crowd.

- **Evidence Extraction and Link Discovery (EELD).*** This project is an effort to identify terrorist groups by developing a suite of technologies to detect patterns between people, organizations, places and things from intelligence messages and law enforcement records, and then use those patterns or links to gather additional information from vast amounts of textual or transactional data including web sites, sensor data, and news reports.

- **Genisys.*** This project is a new effort in 2003 to put together old and new databases so that they can be readily queried. This "ultra-large all-source information repository" could include information about potential terrorists and possible supporters, purchase of terrorist types of material, training and rehearsal activities, potential targets, and status of defenses, as well as research into methods of protecting privacy.[48]

- **Bio-surveillance (re-named Bio-ALIRT IN FY2004):*** This project is an effort to collect and analyze information from non-traditional human, agricultural and animal health data bases in order to develop indicators and models, and set up a prototype bio-surveillance system for a citywide area like Norfolk, Virginia to

---

[47] See Letter from Barbara Simons, Ph.D., and Eugene H. Spafford, Ph.D, Co-Chairs, U.S. Association for Computing Machinery to Senators John Warner and Carl Levin, Senate Armed Services Committee, January 23, 2003; see [www.acm.org/usacm/].

[48] Department of Defense, *FY2003 Budget Estimate, Research, Development, Test and Evaluation, Defense-wide, Volume 1, Defense Advanced Research projects Agency,* February 2002*;* web site address above.

increase DOD's ability to detect a clandestine biological warfare attack.

Machine Translation Projects.

- **Translingual Information Detection, Extraction and Summarization (TIDES).*** TIDES is designed to get critical information quickly for intelligence analysts and operators by developing tools that can rapidly find, summarize, and translate key information in foreign languages.

- **Effective Affordable Reusable Speech-to-Text (EARS):** Anticipated to increase the speed of translation from oral sources by ten to 100-fold (including broadcasts and telephone), as well as extract clues about the identity of speakers, EARS is intended to serve the military, intelligence and law enforcement communities.

- **Multispeaker Environments (MUSE) and Global Autonomous Language Exploitation (GALE):** MUSE and GALE are successor programs to EARS. MUSE is to produce transcripts from command centers and meeting rooms and GALE is to develop techniques for detecting key intelligence in massive amounts of foreign language transmissions.

- **Communicator:** Designed to enable military personnel to get logistical support and tactical information when in the field, prototypes of this "smart phone" have already been deployed on Navy ships.

- **Babylon:** Another battlefield system likely to be deployed in Afghanistan in the next few months, Babylon is intended to aid those in the field by translating foreign phrases for the service member.[49]

Protection of Critical Information Infrastructure.

- **DefenseNet (DNET):** This effort is intended to increase the security and performance of DOD's information infrastructure in handling large volumes of information.

- **Mis-Information Detection and Generation (MIDGET):** A new project in 2003, this effort is designed to detect and reduce DOD's vulnerability to mis-information about adversaries that appears in open-source data.

---

[49] Although Communicator and Babylon are primarily battlefield systems, some elements may be incorporated into the TIA system.

Tools for High-Level Decision Makers.

- **Rapid Analytic Wargaming (RAW):** This project is intended to develop gaming technologies that simulate asymmetric threats to be used by the major commands in training and operational settings.

- **War Gaming the Asymmetric Environment (WAE).*** This effort is an initiative to develop tools and models to help analysts and decision makers predict the behavior and the reactions of terrorists to U.S. actions.

- **GENOA/GENOA II:*** **Project Genoa** attempts to improve collaborative reasoning, estimate plausible futures, and create actionable options among intelligence analysts in various organizations. **Genoa II** seeks to enhance collaboration between people and machines in order to improve support provided by intelligence analysts to policy makers at the military command level, to high level DOD civilian officials, NSA and the Joint Chiefs of Staff for dealing with terrorist threats.

- **Total Information Awareness.*** TIA is to integrate some or all of the efforts above into a prototype system or systems that would create and exploit large-scale, counter-terrorist data bases, develop new analytical techniques and models for mining those data bases so as to improve our ability to detect, anticipate, pre-empt, and respond to terrorist attacks. R&D efforts specifically linked to the TIA system in FY2003 are Human ID at a Distance, EELD, Genisys, Bio-surveillance, TIDES, WAE, Project Genoa and Genoa II, and the TIA integrative effort.