U.S. Department of Homeland Security
Washington, DC 20528

Homeland
Security

April 14, 2008

Ms. Marcia Hofmann
Electronic Frontier Foundation
1875 Connecticut Avenue, N.W.
Suite 650
Washington, D.C. 20009

Re: DHS/OS/PRIV 07-197/Hofmann request

Dear Ms. Hofmann:

This is the final letter in response to your Freedom of Information Act (FOIA) request to the Department of Homeland Security (DHS), dated November 20, 2006 and referred to this office by the Privacy Office on March 7, 2008. We were asked to review 11 documents, consisting of 27 pages, to determine if your requested documents can be released or if the documents are exempt from release.

Of those pages, we have determined that 25 pages of the records are releasable in their entirety, one page is partially releasable and we are withholding one page in their entirety pursuant to Title 5 U.S.C. § 552. I have withheld these documents under FOIA Exemption 5.

> **FOIA Exemption 5** protects from disclosure those inter- or intra-agency documents that are normally privileged in the civil discovery context. The three most frequently invoked privileges are the deliberative process privilege, the attorney work-product privilege, and the attorney-client privilege. After carefully reviewing the responsive documents, I determined that the responsive documents qualify for protection under the deliberative process privilege. The deliberative process privilege protects the integrity of the deliberative or decision-making processes within the agency by exempting from mandatory disclosure opinions, conclusions, and recommendations included within inter-agency or intra-agency memoranda or letters. The release of this internal information would discourage the expression of candid opinions and inhibit the free and frank exchange of information among agency personnel.

You have a right to appeal the above withholding determination. Should you wish to do so, you must send your appeal and a copy of this letter, within 60 days of the date of this letter, to: Associate General Counsel (General Law), U.S. Department of Homeland Security, Washington, D.C. 20528, following the procedures outlined in the DHS regulations at 6 C.F.R. § 5.9. Your envelope and letter should be marked "FOIA Appeal." Copies of the FOIA and DHS regulations are available at www.dhs.gov/foia.

If you need to contact our office again about this matter, please refer to S&T 08-0003.13/Hofmann request. This office can be reached at (202) 254-6819.

Sincerely,

Nicole Marcum
Acting AGC for S+T
Mark E. Rosen
Associate General Counsel for Science & Technology

Enclosure: a/s

ADVISE

Specific Open Questions
March 23, 2007

The following is a list of outstanding questions related to DHS use of ADVISE. Some of these questions have already been asked of various individuals. PRIV would like S&T to confirm the by providing updated/confirmed answers.

In the event S&T would like to refer specific questions to other DHS components, please also send the contact information for that individual to PRIV.

The Interagency Center for Applied Homeland Security Technology (ICAHST)

The following are general questions about the overall capability of ADVISE tools:

1. Please provide a short summary of the results of the testing to date: what works, what does not work? Any overall views, based on tests, for the future of ADVISE?
2. What is the overall value of ADVISE, what can be said about it that fits with the current investment?
3. Can ADVISE identify "unknown" patterns and/or predict?
4. Can ADVISE automatically identify items in data or automatically establish relationships?

   From S&T descriptions, PRIV understands that ADVISE can only work with entities and relationships established by a human – that the advantage of ADVISE is its capability to work with all entities and relationships already established in the data.

   Please clarify S&T's responses in the discussion with Herb Engle from March 12, 2006 – reproduced at the bottom "ICAHST Capability."
5. Does the Ontology create a limit on the data that can be loaded into a deployment of ADVISE technology? What is the interaction between a data load scenario and an ontology if the ontology is more limited than the available data? Is the ontology updated or the data load limited?
6. Has ADVISE ever been used to make any operational decisions?
7. Did any individual or DHS component outside S&T operate an ADVISE pilot?

   S&T reported that S&T operated all the pilots (loading data, identifying relationships, demonstrating results) and that all others only watched.

All-Weapons of Mass Effect (All-WME)

The following are specific questions directly related to the All-WME pilot:

8. A draft Privacy Threshold Analysis states that this pilot originally stared on October 1,2006 and was last updated on October 1, 2006. Please describe the history of this pilot with specific references to those dates.

9. Was any data loaded into the pilot?

   S&T reported that data was loaded from FIBIS (opensource.gov) and CNS (cns.miss.edu).

10. If data was loaded, what are the data elements?

11. If data was loaded, what was the source and range of data?

12. If data was loaded, did any information relate to individuals?

13. Is the pilot intended to relate to groups or individuals?

14. Is the pilot intended to relate to individuals, will the individuals be US Persons?

Remote Thread Alerting System (RTAS)

The following are specific questions directly related to the RTAS pilot:

15. What was the exact start and end dates for this pilot?

    S&T reports the pilot started in 2004 and ended in 2006.

16. What does "decommissioned" mean?

    S&T reports RTAS "decommissioned" at the end of the pilot period.

17. What happened to the data once the pilot ended?

18. Is there a way to determine whether the data supplied actually contained personally identifiable information?

    S&T reports the PIERS data included data fields for name and that the data in this field could be either a business or an individual.

19. Please confirm that there is no personally identifiable information in the census data used with this pilot.

20. During the Intellectual property scenario of the demonstration of RTAS, were any searches conducted regarding the shippers, consignees or notify parties?

    S&T states that the intellectual property search was conducted but that S&T does not recall whether the demonstration searched the details of the shipper, consignee or notify parties.

21. At any point, was the RTAS pilot used to search for information about an individual? (Note question 18 re: the determination of whether personally identifiable information could be included in the data sets.)

ICE Demonstration (ICE Demo)

The following are specific questions directly related to the ICE Demo pilot:

22. What was the exact start and end dates for this pilot? When did the pilot actually start and when it actually end?

    S&T reports the demonstration of the pilot occurred on July 28, 2005.

23. What specific data sources were used in this pilot, what was S&T's source for each data set, and for each data set, what range of data was actually used?

    S&T reports that a small sample of data from up to eight different data sources were provided by ICE to S&T. S&T reports that it does not recall whether data from all of the data sources was used, nor does it recall how much data was loaded from the data sources that were used. Without this specific information it is impossible to identify whether personally identifiable information was in fact used during the pilot and thus it is impossible to accurately determine whether there was a privacy violation.

    The following are the data sets currently identified related to this pilot. For each data set, please:

    - Confirm that it was in fact used;
    - Please describe each data set;
    - The range of data used from each data set;
    - S&T's source for the data set (with specific contact information); and
    - Which System of Records Notice covers each data source.

- SEVIS
- LESC.
- No fly List
- Selectee List
- NORA
- NSEERS
- Unconfirmed Overstays
- SITSDATA

24. Were all the data identified in question 24 were fused (combined) together? Or were different sources fused in different combinations?

25. If the data was combined, did the ontology limit the actual data that was actually combined? If so, what data was actually combined?

Threat Vulnerability Integration System (TVIS)

26. The June 27, 2006 Privacy Threshold Analysis states that this pilot is a new development effort. When did it start and what were the dates of any substantial updates to the pilot?

27. Was all personally identifiable information removed from this pilot?

    S&T reports that all PII was removed from the pilot and all activity related to this pilot has stopped.

28. Please confirm these data source were the actual data sources used. If any other data sources were used please identify those other data sources.

    The following are the data sets currently identified related to this pilot. For each data set, please:

    - Confirm that it was in fact used;
    - Please describe each data set;
    - The range of data used from each data set;
    - S&T's source for the data set (with specific contact information); and

- No Fly List
- Selectee List
- TSC Daily Summaries
- Intelligence Community Message Traffic
- NTIDB
- Patriot Reporting
- SEVIS

ICAHST CAPABILITY

-----Original. Message-----
From: Sand, Peter
Sent: Monday, March 12, 2007 7:59 AM
To: Engle, Herbert
Cc: Hoyt, John; Baicar, Bruce; Jorgensen, Bruce <CTR>
Subject: RE: ADVISE - ICAHST Questions

Herb,

In terms of the actual function of the ADVISE tools (separate from the experience of using it), can you describe what else it can do - in addition to analyzing the relationships between linked nodes?

Does ADVISE have the capacity to identify new patterns itself? Note, this is different from the manually-created scenarios described in the below quote:

"Linking both nodes gives you a pattern for which you can query from.... first we will create a pattern and find it, then we will modify the ontology then find the same pattern again. Results should look like pattern."

Can ADVISE create its own patterns?

Thanks,

Pete
-----Original Message-----
From: Engle, Herbert
Sent: Monday, March 12, 2007 8:26 AM
To: Sand, Peter
Subject: RE: ADVISE - ICAHST Questions

Peter,

If the question is does ADVISE it's self seek out patterns in data and then notify an analyst of a pattern then that would be no. ADVISE is a vary powerful visualization tool. It will graphically depict pattern that are in the data that has been loaded the system but it requires an analyst to identify patterns. An Analyst will query the system asking to see the relationships in the data. ADVISE then take the "raw data" and presents it in the from of links and patterns. The test statement refers to the process of placing a pattern into the system (How A relates to B and how A and B relate to C). Then that pattern is modified and those results are compared to the first pattern. This was part of the Phase 1 testing and was used to verify basic operational functionality.

Herb

-----Original Message-----
From: Sand, Peter
Sent: Monday, March 12, 2007 8:32 AM
To: Engle, Herbert
Subject: RE: ADVISE - ICAHST Questions

Herb,

Just so I am clear, ADVISE as a toolset IS NOT CAPABLE of generating patterns on its own. AN ANALYST must CREATE the LINKS between nodes and a certain SET OF LINKS can be stored as a "PATTERN" and be searched for later and by others.

Did I get it right?

Pete

-----Original Message-----
From: Engle, Herbert
Sent: Monday, March 12, 2007 8:47 AM
To: Sand, Peter
Subject: RE: ADVISE - ICAHST Questions

Peter,

ADVISE can show the link in any data that it has if an analyst creates a query that asks to see that relationship. An analyst does not create a link. He may ask to see what links some node has with another. If you load 1000 data points and ask to see the relationship between all or some of them, then the system will show you. The key is that you can pull data from a number of sources. You might have flight information and phone records. ADVISE will let you see how these two types of information relate by providing a graphical representation of how the data links together.

Herb

All Weapons of Mass Effect (All-WME)

The All Weapons of Mass Effect (All-WME) program is currently housed in the Command, Control and Interoperability Division at the Science and Technology Directorate (S&T). The program assesses the capabilities of foreign and domestic terrorist groups to develop and deploy WME threat agents.

DoE's Lawrence Livermore National Laboratory (LLNL) started the All-WME effort in October 2002, prior to the formation of DHS. S&T supplied its initial funding for All-WME in 2003. In these early stages, All-WME activities relied on existing data management and analysis tools developed by LLNL and Los Alamos National Laboratory (LANL) scientists. One such tool was known as the Knowledge Integration Tool (KIT), and used simple Web-like interfaces. Until 2005, all the WME analysts were exclusively DoE analysts at LLNL and LANL. They analyzed classified message traffic collected by the laboratories' Field Intelligence Elements (FIEs). Such message traffic may include personally identifiable information, that is, data that can potentially identify a person, but does not contain data on U.S. persons.

In FY 2005, S&T began funding an effort to explore whether the ADVISE framework could be used to analyze All-WME message traffic data. An internal test and development capability was set up for that purpose early in 2006. In addition, a limited set of message traffic data was entered into a separate, stand-alone ADVISE framework for performance testing and evaluation. No operational decisions were made from this performance test and evaluation. At this time, there has been no further work on the test and development system.

The ultimate relevance of ADVISE to All-WME includes:

- Capturing information and knowledge from high-value documents that would not be available through other means
- Capturing and sharing knowledge of analysts. For example, analysts may annotate or vet documents or information which should then be shared with other analysts in their organization
- Fusing data from multiple sources or organizations

Development of an ADVISE-based, All-WME pilot, which was initially planned for FY 2007, was halted in 2006 as a result of funding priorities. The pilot would have characterized the capabilities of adversaries by creating a comprehensive and current awareness of WME related materials and illicit trafficking.

Privacy status: The All-WME analysts operate as part of the DoE FIEs. As such, they strictly follow DoE rules for protecting privacy. S&T is currently drafting a PTA to reflect the All-WME initiatives prior to FY2007.

Questions and Answers
For
ADVISE
(Analysis, Dissemination, Visualization, and Semantic Enhancement)

1. **Please thoroughly describe the data-mining tool or activity and the data that is being or will be used.**

   ADVISE is a framework of tools to analyze and visually represent relationships between people, places and events. It is being developed to provide analysts with help in quickly retrieving the right information for their current research and reporting needs. Intelligence analysts depend upon information from a variety of sources such as documents in many different forms: email, database records, web pages, spreadsheets, text files, etc. The volume of information available on a daily basis is tremendous and continues to grow beyond the ability of anyone to read and assimilate all of the data contained. The ADVISE system has two primary capabilities to assist analysts: 1) ADVISE shows relations between entities (people, places, things) from disparate data sources that would otherwise go unnoticed using traditional information retrieval approaches; 2) ADVISE provides a fast, accurate analysis of huge quantities of documents to locate the few that are pertinent to an analyst's current research needs.

   The basic components of ADVISE are a semantic graph, analysis tools, visualization tools, data loading, text processing and documented application programmer interfaces (APIs) between these components. A typical ADVISE deployment consists of rack-mounted servers and support hardware to enable the semantic graph, analysis engines, and text processing. Client workstations execute software that is loaded on demand to query and visualize the data.

   ADVISE is loaded with data selected by the implementing organization per that organization's policies. When data is brought into a functioning ADVISE system, the data loading utilities and text processing utilities extract entities, attributes of entities and relationships between entities from the source documents and data. The extracted entities and relationships are used to construct the semantic graph. Attributes about the entities are stored in an additional data store. A document management system is included to access the documents brought into ADVISE.

2. **Please describe the goals and plans for the use or development of the data-mining tool or activity. For what purpose(s) is the data-mining tool or activity being developed and deployed?**

The goal is to address the growing needs of the intelligence organizations to rapidly sift through and pinpoint the most pertinent and useful documents that analysts can use in the day-to-day compilation of reports about terrorist and proliferation activities.

3. **What is the source of the data used by this data-mining tool or activity?**

The source of data used in ADVISE is determined by the organization that deploys the ADVISE framework (the hardware and software that make up the ADVISE tools). Currently ADVISE has four deployments within DHS; three of these are within the Science and Technology (S&T) directorate and one is in the Office of Intelligence and Analysis (I&A). Of the three S&T deployments, two are operational and use data on chemical, biological, nuclear, and radiological (CBRN) topics. The other S&T deployment is a test and evaluation of the ADVISE tool set. Data loaded in this test system is synthetic or fictional data; there is no correlation with actual people, places, events, or things. The I&A deployment is in pilot development—analysts are being briefed on capabilities, existing I&A data are being test loaded, and evaluation of research utility is being conducted.

4. **For how long has this data-mining tool or activity been in development or operation? If the tool or activity is still under development, what is the target date for its deployment?**

The ADVISE framework has been in development since 2003. The S&T deployments mentioned above have been established since that time. Other DHS components are expected to begin pilot deployments in FY2007 and the I&A deployment is expected to go operational in FY2008.

5. **How effective is the data-mining tool or activity in providing accurate information consistent with the tool or activity's goals?**

The CBRN-related deployments are successfully discovering relationships between entities that would not have been discernable with other query tools and data repositories. These discoveries are providing the basis for reports to agencies that expect to initiate actions based upon these reports. The I&A deployment, because it is still in pilot mode, does not contain sufficient data to provide actionable results. The test and evaluation deployment will complete two phases of testing in FY2007. The results of this testing when completed will provide more concrete, objective answers to this question.

6. **Was a Privacy Impact Assessment conducted prior to the initiation of the data-mining activity or tool? If not, what other assessment has been made regarding the impact of the data-mining tool or activity on privacy or civil liberties generally, and what were its findings?**

Privacy Threshold Assessments have been completed for all deployments. Privacy Impact Assessments have been completed for the I&A deployments.

7. **What are the laws and regulations that govern the information being collected, used, and analyzed by the data-mining tool or activity?**

   Laws considered include, but are not limited to the Electronic Communications Privacy Act, the Wiretap Act, the Pen Register, Trap, and Trace Device Act, the Privacy Act, and the Fourth Amendment to the Constitution, and Executive Order 12333.

8. **What policies and procedures are in place to ensure the security and integrity of the data?**

   Each deployment of ADVISE implements the level of access and security controls appropriate to that organization's policies. ADVISE is secured using a PKI-based security infrastructure that is designed for accreditation at DCID 6/3 PL-3. The security layer includes access control and authentication services to ensure that only individuals who have received approval can access the system and that their access credentials are authentic. Further, ADVISE restricts access to data based upon the role(s) assigned to each individual. ADVISE also implements the concept of communities of interest (COI). The combination of roles and COI enforce the policy that only individuals with approved access to the data at their level of security classification can access the data assigned to their community of interest.

9. **Is privacy and security training required of staff with access to the data?**

   ADVISE depends upon the implementing organization to train its staff on appropriate use of data they load into ADVISE and analyze through the use of ADVISE tools. ADVISE provides capabilities to enforce the privacy and security policies of the implementing organization.

10. **Is access to data limited to those with a need to know?**

    Yes, ADVISE provides capabilities to limit access to data per the policies of the implementing organization to those with a need to know.

11. **What steps are taken to assure the accuracy and integrity of the data?**

    ADVISE is an analysis tool. The implementing organization loads data into ADVISE for analysis and is responsible for assuring the accuracy and integrity of the data per that organization's policies. Once in ADVISE the data is secured using a PKI-based security infrastructure that supports auditing of all user actions.

12. **How is the data secured? Are the databases subject to regular audits?**

    Data contained in ADVISE is secured using a PKI-based security infrastructure that is designed for accreditation at DCID 6/3 PL-3. This infrastructure supports auditing of all user actions. Frequency of audits of ADVISE data is determined by the implementing organization.

What is ADVISE?
The Threat Awareness Portfolio, within DHS Science and Technology, is developing a prototype technology to perform real-time threat analysis and warning to assist in-depth assessments of terrorists' capabilities and intent. The Analysis, Dissemination, Visualization, Insight, and Semantic Enhancement (ADVISE) framework is the center piece of the portfolio's knowledge management tools.

The ADVISE program is a research and development effort comprising three elements:

- Data Science and Data Representation – A flexible computing architecture for collecting, analyzing, and synthesizing threat information from multiple, distributed, and disparate data sources
- Visualization and Analytics - Techniques for visualizing, relating, and synthesizing information of multiple data types and from multiple sources
- Discrete Sciences and Modeling and Simulation - Advanced computing algorithms and hardware architectures for modeling, simulating, and managing threat data in real time and with high resolution

ADVISE will provide a set of core Knowledge Management Tools that enable Information Fusion and Sharing, thus providing the type of capability to "connect the dots" that was described in the *9/11 Commission Report* and described in the *Intelligence Reform and Terrorism Prevention Act of 2004.*

At completion of the research program, ADVISE will deliver a technology or set of technologies, NOT a database OR a data collection activity. It will be applied to maintaining information on terrorists, their capabilities to develop and deploy weapons of mass effect, and their potential activities only.

How does ADVISE work?
The ADVISE semantic graph facilitates assessments by extracting important relationships and correlations from a large amounts of data and producing actionable intelligence. This drives the need for innovative technical solutions that scale with orders of magnitude greater performance than existing analytic capabilities. To draw conclusions from related facts in multiple data sources, DHS must utilize or develop a series of important technologies. The ADVISE approach to modeling information allows new relations ("knowledge") to be discerned in ways that aren't possible with traditional distributed relational-query techniques.

Without such a critical strategic capability, DHS will be unable to identify suspicious linkages and prevent complex asymmetric threats. This research and development effort has produced a prototype that is currently available to analysts at DHS Office of Intelligence and Analysis (I&A). Data currently being loaded into the graph is already extant at I&A.

Legal rights and privacy constraints along with the security of all data sources are protected by this prototype technology. Care has been taken to ensure compliance with respect to privacy, information assurance, and security according to established government policies. The DHS Privacy Office has been engaged in this effort since program conception. Privacy and information assurance technology is based on proven technologies already operational within the intelligence community.

What's Different?

- Privacy and security – Preserve information privacy and security while accomplishing the mission
- Scale – The sheer quantity of data and information
- Dynamic real-time queries – Support real-time updates and queries to the semantic graph
- Mixed domains – Connections between different information domains
- Timely results – Short lifetime for actionable information
- Tracking changes – The temporal nature of information is key to analysis

What is ADVISE?
The Threat Awareness Portfolio, within DHS Science and Technology (S&T), is developing a technology to perform real-time threat analysis and warning to assist in-depth assessments of terrorists' capabilities and intent to develop and employ weapons of mass effect (WME). The Analysis, Dissemination, Visualization, Insight, and Semantic Enhancement (ADVISE) framework is one part of S&T's threat awareness capability.

The ADVISE program is part of a research and development effort comprising three distinct elements: Framework and Architecture, Visualization and Analytics, Discrete Sciences and Modeling and Simulation.[1]

ADVISE will provide a set of knowledge management tools that enable information fusion and sharing across multiple agencies. That need to "connect the dots" was described as critical to terrorism prevention in the *9/11 Commission Report* and was mentioned in the *Intelligence Reform and Terrorism Prevention Act of 2004.*

At the completion of the research program, ADVISE will deliver a technology or set of technologies, NOT a database OR a data collection activity. It will be applied to maintaining information on terrorists; their capabilities to develop and deploy weapons of mass effect; and their potential activities.

How does ADVISE work?
ADVISE is being designed to extract important relationships and correlations from large amounts of data and to produce actionable intelligence on terrorists and WME. This drives the need for innovative technical solutions that scale significantly better than existing analytic capabilities. To draw conclusions from potentially related facts in many, diverse (or heterogeneous) data sources, S&T must utilize or develop a series of new or emerging technologies. The ADVISE approach to modeling information allows new relations ("knowledge") to be discerned in ways that aren't possible with traditional distributed relational-query techniques.[2]

Without such a critical strategic capability, DHS will be unable to identify suspicious linkages and prevent complex asymmetric threats. The research and development effort has produced a prototype that is currently available to analysts at the DHS Office of Intelligence and Analysis (I&A). Data currently being loaded into ADVISE is already extant at I&A.

Legal rights along with the security of all data sources are protected in this prototype technology, as the ADVISE technology includes a security layer. Care has been taken to ensure compliance with respect to privacy, information assurance, and security according to established government policies, and privacy and information assurance technology is based on proven technologies already operational within the intelligence community. The DNI CIO provided funding to make the basic ADVISE infrastructure ICSIS compliant. The security infrastructure protects ADVISE by authenticating of users, providing user credentials to components within ADVISE so that they can enforce application specific access control policies, and by keeping an audit log of requests and responses. The security infrastructure is certified and accredited by DHS at DCID 6/3 Protection Level 3.

What's Different from Current Capabilities?

- Privacy and security – Preserves information privacy and security while accomplishing the mission
- Scale – The sheer quantity of data and information
- Dynamic real-time queries – Supports real-time updates and queries
- Mixed domains – Connections between different information domains
- Timely results – Short lifetime for actionable information
- Tracking changes – The temporal nature of information is key to analysis

Notes

[1] **Architecture and Framework** – This element refers to the ability to ingest, process, and fuse vast quantities of information to find hidden relationships and links, enabling the U.S. to "connect the dots" ahead of time. The objective is a common, flexible, multi-use computing architecture for processing, analyzing, synthesizing, and disseminating massive amounts of threat information from multiple, distributed, and disparate datasets.

Visualization and Analytics - Today intelligence and law enforcement have very large dynamic data sets, but their usefulness is limited by an analyst's ability to understand or comprehend the data. Visualization and analytics seeks to display relationships and the dynamics of vast amounts of intelligence in an intuitive visual way. This element includes techniques for rapidly and easily discovering, relating, and synthesizing diverse information of multiple data types and from multiple sources, including information extraction and pattern discovery from massive, diverse data sets.

Discrete Sciences and Modeling and Simulation - Our ability to understand and simulate the dynamics of a major event, like Hurricane Katrina, is limited by our modeling and simulation technologies. In-depth discrete imulation will enable us to predict in real-time major event consequences for scenario planning, and incident response. This element includes advanced computing algorithms and hardware architectures for modeling, simulating, and managing threat data in real time and with high resolution for accurate, timely threat assessment.

[2] Known as a "semantic graph", which portrays the relationships between the various data elements.

What is ADVISE?
The Department of Homeland Security, Science and Technology Directorate, is developing a prototype semantic graph-based technology to perform real-time threat analysis and warning to assist in-depth assessments of data. The Analysis, Dissemination, Visualization, and Semantic Enhancement (ADVISE) framework is a framework of R&D knowledge management tools. ADVISE is not an operational system, and ADVISE does not collect data. It is a set of R&D information technology tools being developed to aid human analysis of large amounts of already collected data.

The ADVISE R&D program is a research and development effort comprising three elements:

- Data Science and Data Representation – A flexible computing architecture for collecting, analyzing, and synthesizing threat information from multiple, distributed, and disparate data sources
- Visualization and Analytics - Techniques for visualizing, relating, and synthesizing information of multiple data types and from multiple sources
- Discrete Sciences and Modeling and Simulation - Advanced computing algorithms and hardware architectures for modeling, simulating, and managing threat data in real time and with high resolution

At completion of the research program, ADVISE is intended to deliver a technology or set of technologies, NOT a database OR a data collection activity.

How does ADVISE work?
The semantic graph facilitates assessments by extracting important relationships and correlations from a plethora of data and producing actionable information. ADVISE performs data fusion across large disparate data sets. This drives the need for innovative technical solutions that scale with orders of magnitude greater performance than existing analytic capabilities. Without such a critical strategic capability, we will be unable to identify suspicious linkages and prevent complex asymmetric threats.

Legal rights and privacy constraints along with the security of all data sources are protected by this prototype technology. Care has been taken to ensure compliance with respect to privacy, information assurance, and security according to established government policies in any pilot programs. The DHS Privacy Office has been engaged in this effort since program conception. Work is progressing on how IT tools can be assessed for privacy, before specific data sets are identified. Currently, privacy can only be assessed after a system to be developed (and its data) are defined.

What's Different?

- Privacy and security – Preserve information privacy and security while accomplishing the mission
- Scale – The sheer quantity of data and information
- Dynamic real-time queries – Support real-time updates and queries to the semantic graph

TVIS Pilots Data Sources

Database	Type of Information	Source
TSA No Fly List	One time load of the spreadsheet. This source almost certainly contains US Person data of individuals meeting the criteria to be on the list. Future plans call for automatic updates.	The I&A TSAwatchlist email distribution list. A sample of this data has been loaded, but there is no automatic load process in place.
TSA Selectee List	One time load of the spreadsheet. This source almost certainly contains US Person data of individuals meeting the criteria to be on the list. Future plans call for automatic updates.	The I&A TSAwatchlist email distribution list. A sample of this data has been loaded, but there is no automatic load process in place.
TSC Daily Summaries	One time load of the daily summaries of reported incidents from the Terrorist Screening Center. This source can contain US Person information associated with an incident. Future plans call for automatic updates.	The Terrorist Screen Center email daily distribution of encounters. A sample of this data has been loaded, but there is no automatic load process in place. POC of contact was Ben Stefano.
IC Message Traffic	One time load of manually tagged high-interest data from unstructured text message traffic. This source may contain US Person information. Future plans call for semi-automated ingestion of this data.	Email postings made on classified network from members of the operations center. Approximately 200 documents were manually entered.
NTIDB	One time load of I&A's National Threat Incident Database, which aggregates data from other sources. This source does contain US Person information. Future plans call for automatic	A computer-to-computer connection to the NTIDB system

	updates.	
Patriot Reporting	One time load of unstructured text from select entries from the US Patriot Reports. This source does contain US Person information. Future plans call for automatic updates.	Spreadsheet summary of patriot reports supplied by James Szrama (HITRAC analyst)
SEVIS	One time load of a portion of entries from SEVIS—the loaded records are foreign exchange students who are in the US studying for an advanced degree—This source does not contain US Person information	Spreadsheet summary of SEVIS records supplied by Scott Summey

SUMMARY DESCRIPTION OF TVIS USE OF DATA SETS

DATA

1. NTIDB
   - Elements:
   - Range:
     - all data from NTIDB inception until approximately June of 2005
   - Purpose:
     - this data set was loaded primarily to demonstrate data fusion. If an individual's name was identified through the link analysis, then the fact that the individual was on the No Fly list would be evident because a link to the no fly list would be included in the list of documents available for review by the analyst.
2. Selectee
   - Elements:
     - Name of individual
   - Range:
     - all data as of 9/11/2006
   - Purpose:
3. No Fly
   - Elements:
   - Range:
   - Purpose:
   - all data as of 9/11/2006
   - Purpose:
     - this data set was loaded primarily to demonstrate data fusion. If an individual's name was identified through the link analysis, then the fact that the individual was on the No Fly list would be evident because a link to the no fly list would be included in the list of documents available for review by the analyst.
4. Daily Summaries
   - Elements:
   - Range:
   - Purpose:
   - All data from 1/1/2006 to 3/31/2006 and

- One day in November 2006 (November 3, 2006--no particular reason for this date, it was just first day in November for which we were starting to get each day's worth of reporting rather than getting it for an entire calendar quarter).
- Purpose:

5. Patriot Reporting

- Elements:
- Range:
- Purpose:
- Last half of 2005, first quarter of 2006)
- Purpose:

6. SEVIS

- Elements:
- Range:
- Purpose:
- Foreign Students enrolling in graduate school in the Fall of 2006
- Purpose:

7. Message Traffic

- Elements:
- Range:
- Purpose:
- This data was originally entered manually from unstructured text messages. The data was first loaded between April and June of 2006. The data was too hard to map into TVIS and no actual demonstration occurred.
- Purpose:

DATA IN COMBINATION

- NTIDB + No Fly + Selectee + Daily Summaries
- No Fly + Selectee + Daily Summaries + Message Traffic + Patriot Reporting
- No Fly + Selectee + Daily Summaries
- SEVIS

ADVISE Pilots Data Sources
DHS S&T Threat Awareness Portfolio

Note: All databases and data extant on analysts' desktops.

ADVISE Pilots	Database	Type of Information
Threat-Vulnerability Integration System at DHS Office of Intelligence and Analysis	TSA No Fly List	One time load of the spreadsheet. This source almost certainly contains US Person data of individuals meeting the criteria to be on the list. Future plans call for automatic updates.
	TSA Selectee List	One time load of the spreadsheet. This source almost certainly contains US Person data of individuals meeting the criteria to be on the list. Future plans call for automatic updates.
	TSC Daily Summaries	One time load of the daily summaries of reported incidents from the Terrorist Screening Center. This source can contain US Person information associated with an incident. Future plans call for automatic updates.
	AMHS--IC Message Traffic	One time load of manually tagged high-interest data from unstructured text message traffic. This source may contain US Person information. Future plans call for semi-automated ingestion of this data.
	NTIDB	One time load of I&A's National Threat Incident Database, which aggregates data from other sources. This source does contain US Person information. Future plans call for automatic updates.
	Patriot Reporting	One time load of unstructured text from select entries from the US Patriot Reports. This source does contain US Person information. Future plans call for automatic updates.
	SEVIS	Portions of the Student and Exchange Visitor Information System database from ICE. This source contains person information for foreign students and dependents. Future plans call for additional loads of other segments of SEVIS.
ICE Demonstration (short term demo – ended in FY06)	TSA No Fly List	One time load of the spreadsheet. This source almost certainly contains US Person data of individuals meeting the criteria to be on the list. Future plans call for automatic updates.
	TSA Selectee List	One time load of the spreadsheet. This source almost certainly contains US Person data of individuals meeting the criteria to be on the list. Future plans call for automatic updates.

	NSEERS	One time load of the National Security Entry Exit Registration System database from ICE. This source contains non-US person information for aliens. Future plans call for automatic updates.
	SEVIS	One time load of the Student and Exchange Visitor Information System database from ICE. This source contains person information for foreign students and dependents. Future plans call for automatic updates.
	NORA Data Sample	A one time load of a small subset of ICE in a format just before loading into ICEPIC. This source can contain US Person information. Future plans call for automatic updates.
	Unconfirmed Overstay List	A one time load over a few days of persons who have overstayed their visa. This source contains non-US person information for aliens. Future plans call for automatic updates.
	LESC Frontline	A one time load of the ICE Law Enforcement Support Center's Frontline database for a few months. This database tracks queries from Law Enforcement to ICE. This source may contain US Person information. Future plans call for automatic updates.
[redacted]	[redacted]	[redacted]
	Promed Mail	One time load of the publicly available Federation of American Scientists initiative for global monitoring of emerging diseases containing email alerts of outbreaks and reports. Can contain US Person information as researchers and organizations. Future plans call for automatic updates.
	OIE Incident Reports	One time load of a list of disease outbreaks by country from World Organization for Animal Health. This source has no US Person information. Future plans call for automatic updates.
	OIE: Reference Experts and Laboratories	One time load of a publicly available list of experts and laboratories outbreaks by country from World Organization for Animal Health. This source has US Person information

		automatic updates.
	US Army SBCCOM	Soon to be loaded information about various pathogens including lethality (historical, since the data source has disappeared.) This source has no US Person information. Future plans call for automatic updates.
All Weapons of Mass Effect ADVISE pilot at Lawrence Livermore National Laboratory	IC Message Traffic	One time load of manually tagged high-interest data from unstructured text message traffic. This source may contain US Person information. Future plans call for semi-automated ingestion of this data.
	FBIS	One time load of manually tagged high-interest data from unstructured text from the Foreign Broadcast Information Service. This source is likely to contain US Person information as reported. Future plans call for semi-automated ingestion of this data.
	CNS	One time load of manually tagged high-interest data from unstructured text Center for Non-proliferation Studies at the Monterey Institute. This source is likely to contain US Person information. Future plans call for semi-automated ingestion of this data.

ADVISE PILOTS

Summary

CURRENT PILOTS - ALL STOP ORDER FROM OGC

These MAY be known outside DHS

1. All-Weapons of Mass Effect (All-WME) to assesses chemical, biological, radiological, nuclear, cyber, and advanced explosives threats as well as threats against infrastructure to understand terrorist group capabilities and intent in each area with the goal of developing an overall understanding of the types of threats and tactics likely to be used. *S&T states no PII; PRIV waiting for S&T confirmation.* PTA & PIA both in draft. S&T work for I&A

2. Threat Vulnerability Integration System (TVIS) to identify helpful opportunities to test the capability of ADVISE technology to help analysts in I&A. – *All data usage and result data are covered by existing SORNs. PTA complete. PIA in draft* .S&T work for I&A

3. Biodefense Knowledge Management System (BKMS) to help DHS analyze and characterize biological threats posed by terrorists. – *S&T says: Not using ADVISE technology yet; may be in the future.* PTA & PIA both in draft. *Note: Recent discussions indicate separate development effort using ADVISE; requires NEW PTA* .S&T work for I&A, Chief Medical Officer

THIS IS NOT A PILOT, IT IS A TEST FACILITY TESTING ADVISE - AND OTHER TECHNOLOGY

This is NOT a Pilot

4. The Interagency Center for Applied Homeland Security Technology (ICAHST) is a research facility managed by S&T to test technologies like ADVISE for capabilities and performance – *PTA conducted before testing with data started and determined no PIA required: Not using PII. PRIV co-chairing Civil Liberties Working Group (CLWG) to oversee all ICAHST research – with Privacy and Civil Liberties Office of the Office of the Director of National Intelligence.* PTA complete, no PIA necessary. Pure S&T

THESE ARE PILOTS WE IDENTIFIED DURING OUR REVIEW - THEY BOTH ENDED IN 2006

These are NOT known outside DHS

5. Remote Thread Alerting System (RTAS) to identify anomalous shipments based on the cargo type and originating country. – *S&T reports no recollection of PII. The data could contain PII if a person's name was used instead of a business name. According to S&T, no retrieval by personal identifier.* Ended in 2006. S&T work for CBP

6. ICE Demonstration (ICE Demo) to determine whether ADVISE would help ICE make better use of the data it receives. – *S&T reports it does not recall which data sets were actually used and how much of any of the data were actually used. PRIV is waiting for more information re: regarding the details. There was no retrieval by personal identifier.* Ended in 2006. S&T work for ICE

From: Tom Bates [twbates@llnl.gov]
Sent: Friday, December 01, 2006 1:21 AM
To: Sand, Peter; Kim Minuzzo; Shepherd, David; Landsberg, Alexandra (Sandy)
Cc: Marcson, Nicole
Subject: Re: PIA needed for BKC

Hi Peter,

The BKC is headquartered at LLNL, which is a DOE GOCO. Our funding from DHS is received/negotiated annually on a project basis. We will not be part of the soon-to-be announced DHS FFRDC that will manage the other parts of NBACC (the BTCC and the NBFAC). Does that influence the ability to demo the tools to DHS customers before the PIA is completed? As you are aware, analysis of openly available, published scientific information is the cornerstone of our system. Removing names from the publicly-available data sources renders the system completely useless. If it helps, the names we have are from scientific pubs and typically only contain first initial and last name, which are poor personal identifiers, particularly for common names.

We will complete the PIA asap. It would be helpful if you could provide one or more already completed/accepted PIA from a similar project to understand of the level of detail required and possibly text that could be re-used. Have any of the other ADVISE-related tools/systems completed this process?

Kind regards,

Tom

At 05:59 PM 11/30/2006, Sand, Peter wrote:
>All, (note: also cc'ing Nicole Marcson - OGC for S&T)
>
>Based on subsequent discussions I now understand that that LLNL is
>functionally part of DHS, as an FFRDC - someone please correct me if I'm
>still off base on this.
>
>Given that even the pilot is being conducted by DHS, DHS is responsible
>for providing privacy protections - which at this point means that:
>
>1. A Privacy Impact Assessment must be completed prior to the use of
>personally identifiable information; and also that
>
>2. A System of Records Notice may also be required based on what appears
>to be the retrieval of personal information by personal identifier - the
>author's names. The final determination of the requirement to draft a
>SORN will be based on the results of the Privacy Impact Assessment.
>
>** Most importantly ** the personally identifiable information that is
>in the pilot should be removed or otherwise ceased to be used until at
>least the PIA and possibly the SORN are finalized.
>
>This is particularly important if you would want to show this pilot to
>others - all Personally Identifiable Information must be removed until
>the privacy compliance documentation is finalized.
>
>One way to work around this would be to substitute synthetic data (made
>up names) for the actual names. This way you can still demonstrate the
>functionality without using real data about real people.
>
>Please let me know if you have any questions and if you would like help

>drafting any of the privacy compliance documentation.
>
>Thanks,
>
>Pete
>------
>Peter E. Sand, J.D., CIPP/G
>Director of Privacy Technology
>The Privacy Office
>U.S. Department of Homeland Security
>Washington, DC 20528
>Tel: 571-227-4134
>Fax: 571-227-4171
>peter.sand@dhs.gov
>www.dhs.gov/privacy
>------

Tom Bates, Ph.D.
DHS Biodefense Knowledge Center
Lawrence Livermore National Laboratory
7000 East Ave, L-179
Livermore, CA 94550
twbates@llnl.gov
Voice: 925 423-3055
Pager: 888 415-1428
Cell: 925 337-1215
Fax: 925 423-6435