



Privacy Impact Assessment
for the

Group Violent Intent Modeling Program

April 25, 2008

Contact Point

Dr. Jennifer O'Connor
Science Advisor, Human Factors Division
Science and Technology Directorate
Department of Homeland Security
202.254.6716

Reviewing Official

Hugo Teufel III
Chief Privacy Officer
Department of Homeland Security
(703) 235-0780



Abstract

This Privacy Impact Assessment describes the research and development objectives of the Department of Homeland Security (DHS) Science and Technology (S&T) Directorate's Human Factors Division Group Violent Intent Modeling (GVIM) project. The goal of GVIM is to determine whether including social and behavioral theories and concepts (from established research) in a software tool that is used to analyze group behaviors and motivations will improve the ability of analysts to identify indicators that could predict group violence. The project will develop a social and behaviorally based framework of theories and concepts that includes modeling and simulation tools to improve the efficiency and accuracy of intelligence analysts examining the likelihood of a group choosing violence to achieve its goals. This PIA is necessary because personally identifiable information will be collected as part of the research and development effort.

Overview

The Human Factors Division (HFD) in the Science and Technology Directorate is charged with researching, developing, and enhancing the analytical capability of DHS to understand the motivation, intent, and behavior of violent groups so it can better prevent group-driven terrorist attacks. There is a significant amount of research available concerning the motivations of individual terrorists, but significantly less information regarding the motivations and behaviors of groups likely to engage in violent activities including terrorism. HFD is funding GVIM to better understand violent groups and to use that understanding to develop an analytical process that DHS can use to improve its abilities to prevent terrorist attacks.

The purpose of GVIM is to develop an analytical tool that uses modeling and simulation to interpret the motivations and behaviors of violent groups and identify indicators that could predict when a group will engage in violent activity, thereby mitigating threats and hazards to the nation. The goal of the GVIM project is to improve the efficiency and effectiveness of DHS's analytical process and tools to better support the Department's intelligence analysts.

The task of gathering and integrating group-focused social and behavioral research involving group violence is significant because it requires surveying across multiple disciplines including sociology, psychology, political science, and anthropology. GVIM seeks to go beyond a simple compilation of known facts about group violence. GVIM will distill academic literature and publicly available information about groups (both violent groups and non-violent groups) to a set of specific criteria that can be used to create an analytical tool based on structured metrics that would provide a more sophisticated view of the likelihood of group violence including terrorism.

Throughout this document, the term "framework" refers to the collection of social and behavioral theories and concepts that are used to study, analyze, and predict group behavior. The term "system" refers to the integration of the framework into a software platform. The term "tool" refers to the intended use of GVIM to support intelligence analysts.



Three groups will participate in the GVIM project:

1. The Research Team. Department of Energy National Laboratories (Pacific Northwest National Laboratory, Oak Ridge National Laboratory, and Los Alamos National Laboratory) referred to collectively as the “research team.” The research team will gather the detailed data and use it to develop the GVIM framework, system, tools, and aggregated results that I&A will evaluate during demonstrations of the final results of the GVIM project.
2. The DHS Intelligence & Analysis Directorate (I&A). I&A is the intended customer of the GVIM project. I&A will establish the criteria used to determine whether GVIM would improve I&A’s analytical process. I&A will not view or use any of the detailed data within GVIM and only see the aggregated results of the GVIM project – with two exceptions.
 - The first instance in which I&A may access detailed data is if I&A contributes a new set of data to the research team for use in GVIM. In the event that I&A does contribute detailed data, the research team will treat the contributed data the same as it treats the data it obtains through its own research and may include aggregations of that data as part of the GVIM demonstration back to I&A.
 - The second instance in which I&A may access detailed data is if, during a demonstration of GVIM, I&A asks a question that cannot be answered using the aggregated results, the research team may reveal a select amount of the detailed data. In these specific situations, the research team will reveal only the minimal amount needed to answer the specific question and then will make every effort to avoid revealing any personally identifiable information that might be contained in the detailed data.

Regardless of I&A’s questions, neither the research team nor S&T will use GVIM to search or present information about individuals. While the detailed data might support a search targeted to individuals, neither GVIM nor any other capability associated with GVIM, will not be used to search for individual names or other identifying characteristics. In no circumstances will the research team or any component of DHS search for PII or use PII as part of any search criteria for any data (including PII) contained in any data set used as part of GVIM. In addition, there will be no attempt to use GVIM or the data supporting GVIM to learn more about individuals from the data used as part of this research effort. The entire and only use of the data and technology associated with the GVIM project will focus solely on groups, not on individuals.

3. S&T HFD. The DHS Science & Technology Directorate (S&T), Human Factors Division (HFD). S&T HFD will provide funding and general oversight of the GVIM project. S&T will not view or use any of the data contained in the GVIM system and will only see aggregated data in the form of overall results and general reports.

Based upon its research of academic literature and other publicly available materials, the research team will demonstrate GVIM’s capabilities to I&A. The demonstrations for I&A will present GVIM’s capability to more easily analyze the likelihood that a group will engage in violent behavior. Through the



demonstrations, I&A will determine whether GVIM would be an effective tool to support I&A's analytical requirements. These demonstrations will be for research and development (R&D) purposes only and will not be used by DHS to make or support any operational decisions. The research team will use I&A's assessment to further refine GVIM.

Section 1.0 Characterization of the Information

The following questions are intended to define the scope of the information requested and/or collected as well as reasons for its collection as part of the program, system, rule, or technology being developed.

1.1 What information is collected, used, disseminated, or maintained in the system?

The research team will collect three categories of information:

1. Social and behavioral science theories and concepts. These theories and concepts are used to analyze information about groups, including their motivations and behaviors.
2. Factual information about groups from public sources such as newspapers, scholarly journals, and publicly available databases. Groups will be selected based on the availability of open source information about the group and on ideological statements of the group. For example, information may be gathered about a group stating publicly that the group will use violence to achieve its goals. The type of information included will focus on factors generally related to groups (e.g., type of group, group size, history of violence, nature of violent activities, targets of violent activities, stated ideology, organizational structure).
3. Personally identifiable information (PII) about individual members of those groups identified in the public source information discussed above. If the publicly available information about the groups also includes information about individual members of those group, the research team may use that PII to further inform the analysis of the groups. Examples of the type of PII that may be used may include individual's name and group affiliation, age, level of education, job history, and birthplace. The specific PII that would actually be used will only be known when the research team actually researches the publicly available information about the groups. Regardless of the specifics, however, the research team will only use PII that incidentally appears in the articles about the groups. The research team will not conduct specific research about individuals and neither will any component of DHS as part of GVIM. Only PII found incidentally would be used and then only to the extent the PII provides greater insight into characteristics of groups.

GVIM is a research project and as the relevant theory and research are integrated into the system's framework, the set of data collected will be refined. As part of the refining process, some or all of the PII may be eliminated from the database as research and development advances. A successful analytic framework provides parsimony in terms of what data is used



and how it is organized for the analysts. In most cases, it is the aggregation of data which provides indicators, not the detailed information originally researched.

As discussed in section 1.5, below, the research team will use Subject Matter Experts to review any PII that is identified and these experts will apply their professional experience and expertise to evaluate the credibility and value of any PII that is used in GVIM.

PII may assist the research team in identifying a particular group's size, how long a particular group existed, and potentially to identify other characteristics of a single group or shared characteristics across to multiple groups that could only be identified through the characteristic of individual group members. In addition, PII may be used to identify a particular group even though the group may have changed names over time – identifying consistent individual members would support a conclusion that different group names point to a single group.

1.2 What are the sources of the information in the system?

The research team will use information only received from open sources to include newspapers, scholarly journals and publicly available resources. In certain instances as part of the demonstration, I&A may add additional open source information. During the demonstrations, I&A analysts will be evaluating the usefulness of GVIM in order to test the capabilities of the tool. In order to facilitate this evaluation, I&A may provide supplemental information. In all cases, whether the data used in GVIM is obtained directly by the research team or provided by I&A, the data will always come from publicly available sources and not from operational DHS systems. In addition, all the data will be historical in nature (i.e., more than a year old).

Should any of the data that I&A proposes to include in the demo fall outside this PIA's description in Section 1.1, this PIA will be amended or a new PIA will be submitted.

1.3 Why is the information being collected, used, disseminated, or maintained?

All information, including PII, is being collected, used, and maintained to facilitate research and development of an analytical framework designed to estimate whether a group will engage in violent activities. The information will be used in the development and testing of models and simulations of emerging group violent behaviors, and to demonstrate GVIM's capabilities to I&A. PII will only be used if it demonstrates value in better understanding a group's characteristics as related to the group's violent activities.



1.4 How is the information collected?

Information, including PII, will be collected directly by the research team from publicly available sources both printed and online. As indicated above, I&A may also contribute additional publicly available information to the research team. The research team will also use a search engine that utilizes keywords to search the World Wide Web and other publicly available information resources for information about groups. This tool will collect data which will be used to test the GVIM framework. The tool is not an automated extraction tool, but relies upon pre-defined keywords in order to locate information.

Researchers will not search for information about individuals, only information about groups. If individuals are named in the articles about the groups, that additional information will be used to provide greater insight into the groups.

S&T will not collect nor will it be provided with the detailed data collected for GVIM.

1.5 How will the information be checked for accuracy?

Subject matter experts (SMEs) from the academic community will review the information for accuracy. SMEs will be experts in information technology, group behavior, terrorism analysis, and/or modeling. SMEs will have access to provenance information (i.e., the specific source of a particular document) for each piece of data and will provide credibility ratings regarding the value of the source and thus the information in the source article, during the modeling process for data based on their knowledge of the data sources. For example, an SME familiar with a particular newspaper and its regular contributing reporters would be able to assess the credibility of an article from that newspaper written by a particular reporter. In addition, multiple sources will be consulted for redundancy.

1.6 What specific legal authorities, arrangements, and/or agreements defined the collection of information?

The Homeland Security Act of 2002 [Public Law 1007-296, §302(4)] authorizes the Science and Technology Directorate to conduct “basic and applied research, development, demonstration, testing, and evaluation activities that are relevant to any or all elements of the Department, through both intramural and extramural programs.” In exercising its responsibility under the Homeland Security Act, S&T is authorized to collect information, as appropriate, to support R&D related to improving the security of the homeland.

1.7 Privacy Impact Analysis: Given the amount and type of data collected, discuss the privacy risks identified and how they were mitigated.

The potential risk to the individual would be inappropriately associating an individual with a violent group. This risk is mitigated by the fact that the project is a research and development effort and information collected for this research will not be used by DHS to make operational decisions. The names



and other information in the data set will not be used to make determinations about individuals or to seek additional information about individuals. Except for limited circumstances, the detailed data, including any collected PII, will only be available to the research team, and then only used to develop the GVIM framework and tool. All information will be collected from the publicly available sources. The data repository is kept on a firewall protected network to protect the information from unauthorized access. At the conclusion of the project, all data will be destroyed.

If information about individuals is contained in the public information about particular groups, that individual information will be used only to identify additional information about the groups. As noted above, a group may change names over time. If publicly available information indicates that membership is consistent across multiple-named groups, it may help the research team to build a better framework and thus a better tool for I&A.

Section 2.0 Uses of the Information

The following questions are intended to delineate clearly the use of information and the accuracy of the data being used.

2.1 Describe all the uses of information.

Information will be used to develop and test an analytical tool that uses modeling and simulation to interpret the motivations and behaviors of groups and identify indicators that could predict when a group will engage in violent activity. Specifically, the information will be used to assess the effectiveness of identifying factors that contribute to whether a group will engage in violence. As the data used in GVIM during S&T's research and development is historical (i.e., the events occurred in the past and the outcome is already known), S&T will measure GVIM's effectiveness by comparing the system's assessment of a group's likelihood to engage in violence to the actual known history of the group. The information will not be used to make or support operational decisions.

The research team will use the research data to produce a list of factors found to be relevant for prediction of group violence. Factors consistently associated with group violence will be considered "relevant for prediction."

GVIM will enable analysts to test their hypotheses about group behavior. The research team is building GVIM as an analytical tool that will use modeling and simulation to interpret the motivations and behaviors of groups and identify indicators that could predict when a group will engage in violent activity including terrorism. The operational version of GVIM will not replace analysts, but will support analysts by highlighting these same factors found while analyzing groups.

Except in the limited situations described above, the research team will be the only users of detailed information and thus the only users of PII that might be collected as part of GVIM. As mentioned above, I&A may suggest the research team add additional publicly available information to the GVIM data set. In those isolated situations, the research team will treat the recommended data the same way it treated the data the research team identified on its own and will only present aggregated data back to I&A during the demonstrations.



The research team will not search or retrieve information by an individual's name or other personal identifiers. Information will be organized and used according to groups and not by information related to individuals. The research team will not conduct any directed research of individuals and will only use PII found while researching groups to learn more about groups, not about individuals.

2.2 What types of tools are used to analyze data and what type of data may be produced?

The tools used to analyze the data include searches by key word and user defined content analysis tools, modeling and simulation tools, and the opinions of subject matter experts.

2.3 If the system uses commercial or publicly available data please explain why and how it is used.

The system will use data from the publicly available source such as newspapers and journals. As discussed above, incidental PII may be found in articles about groups and if the research team and the research team's Subject Matter Experts determine the PII would further the understanding about the groups, that PII may be used in GVIM but only to present information about groups, never about the individuals who may be mentioned in the research data..

2.4 Privacy Impact Analysis: Describe any types of controls that may be in place to ensure that information is handled in accordance with the above described uses.

GVIM is a research project. Except for the limited situations described above, only the research team will access PII and then, only to better understand group behavior. The research team will not research individuals and will only organize and use the detailed data by reference to groups, not individuals. I&A will evaluate the system based on the degree to which using the system improves I&A's analytic process. Should I&A wish to acquire GVIM and use it to support operational decision-making, GVIM would be transitioned to I&A and a new PIA would be completed.

Section 3.0 Retention

The following questions are intended to outline how long information will be retained after the initial collection.

3.1 What information is retained?

During the life of the GVIM project, all source data including PII that might be included in the information about groups, will be retained, as well as any aggregated information used to generate the overarching model factors that will be used to predict group violence. At the conclusion of the GVIM project, all source documents will be destroyed. Only the overarching factors found to predict group violence will be retained; this information will not contain PII.



3.2 How long is information retained?

Until an approved records retention and disposal schedule is established, records will be retained throughout the research and development process, including the potential development of “GVIM version 2.0” based on the results of initial GVIM demonstrations. S&T is currently working with DHS headquarters to determine an appropriate retention schedule. Upon receiving approval, S&T will dispose of records according to the received instructions. Currently, S&T plans to destroy all source documents at the conclusion of the GVIM project.

3.3 Has the retention schedule been approved by the component records officer and the National Archives and Records Administration (NARA)?

DHS S&T is currently working with DHS headquarters to determine an appropriate retention schedule. After DHS headquarters proposes the schedule it will be submitted to NARA for final approval.

3.4 Privacy Impact Analysis: Please discuss the risks associated with the length of time data is retained and how those risks are mitigated.

To mitigate any risks associated with data retention, the data will be destroyed at the conclusion of the project. While GVIM is an S&T research project, the data will not be used for any purpose other than the refinement and development of the GVIM framework. Should I&A wish to acquire GVIM and use it to support operational decision-making, the system would be transitioned to I&A and a new PIA would be completed.

Section 4.0 Internal Sharing and Disclosure

The following questions are intended to define the scope of sharing within the Department of Homeland Security.

4.1 With which internal organization(s) is the information shared, what information is shared and for what purpose?

Selected analysts from I&A will have access to GVIM to test the efficiency and effectiveness and to provide feedback to direct future development of GVIM. GVIM will include both the framework of concepts and theories that enables analysis of the data and the data set. Depending on the research question being explored in a given demonstration, PII may or may not be included in the data set. (The PII enables researchers to maintain the accuracy and consistency of the data. For example, the name of a group might change several times during a ten-year period while the name of the group’s leader stayed the same.) All analysts will complete basic privacy training, and the information will not be used operationally.



4.2 How is the information transmitted or disclosed?

During the research and development phase, information will be aggregated and stored on up to three stand alone computers. PII will only be disclosed to the GVIM research team and those I&A analysts who may be involved in demonstrations of GVIM that specifically require PII. All access to PII within GVIM will be granted based on an established need to know for the purpose of the demonstration projects.

4.3 Privacy Impact Analysis: Considering the extent of internal information sharing, discuss the privacy risks associated with the sharing and how they were mitigated.

The risk of internal information sharing is that the sharing could result in an adverse consequence for an individual. To mitigate that risk, the information will only be internally shared with the research team or I&A research analysts for R&D purposes, and will not be used for any operational purpose. Security measures include controlled access and system firewalls.

If I&A chooses to acquire GVIM, GVIM will be migrated to a client server environment after going through Certification & Accreditation. I&A would submit a PIA for the system prior to engaging in operational use. During the R&D phase, no PII data is transmitted outside the GVIM framework. PII will only be disclosed to the research team and I&A analysts working on the GVIM project who have an established need to know.

Section 5.0 External Sharing and Disclosure

The following questions are intended to define the content, scope, and authority for information sharing external to DHS which includes Federal, state and local government, and the private sector.

5.1 With which external organization(s) is the information shared, what information is shared, and for what purpose?

The information is shared with the research team, which includes authorized personnel at the Department of Energy's laboratories: Pacific Northwest National Laboratory, Oak Ridge National Laboratory and Los Alamos National Laboratory. Should the GVIM project be discussed with any personnel external to DHS or the research team, only group-level information (and no names or other PII) will be disclosed.



5.2 Is the sharing of personally identifiable information outside the Department compatible with the original collection? If so, is it covered by an appropriate routine use in a SORN? If so, please describe. If not, please describe under what legal mechanism the program or system is allowed to share the personally identifiable information outside of DHS.

The only external access to the information will be the research team from the National Labs and contractors hired by DHS to work specifically on this project. The source data for GVIM comes largely from public news reports and reports generated by nonprofit organizations. The information was originally collected by those public news or academic organizations with the intent of educating the public and enhancing public safety with respect to group violence. GVIM's use of the source data is intended to facilitate the development of an analytical framework that can be used by analysts to enhance analytic abilities, which is compatible with the original intent of enhancing public safety. The Intelligence Reform Act authorizes DHS to work with other agencies and the intelligence community to protect the homeland. A System of Records Notice is not required for this effort, as the data collected are not specifically about an individual but are about groups and the information will not be retrieved by a personal identifier.

5.3 How is the information shared outside the Department and what security measures safeguard its transmission?

Except in those limited situations described above, only aggregated information and analytical factors will be shared with the Department. Each Lab stores the information in a password-protected system protected by a firewall.

5.4 Privacy Impact Analysis: Given the external sharing, explain the privacy risks identified and describe how they were mitigated.

The risk of external information sharing is that the sharing could result in an adverse consequence for an individual. To mitigate that risk, only the research team will gather and use the detailed data that might contain PII. In those limited situations where PII may be revealed in a demonstration, the data will only be used to evaluate GVIM's capability to assist I&A analysts and will not be used for any operational purpose. Security measures include controlled access and system firewalls.

Section 6.0 Notice

The following questions are directed at notice to the individual of the scope of information collected, the right to consent to uses of said information, and the right to decline to provide information.



6.1 Was notice provided to the individual prior to collection of information?

This PIA serves as notice regarding how DHS will use PII that may be found in the publicly available data used to develop GVIM. General notice of the existence of the information is available to the individual by searching the same publicly available data sources that the research team intends to search.

6.2 Do individuals have the opportunity and/or right to decline to provide information?

Not with DHS. All information is collected from publicly available sources such as newspapers, academic journal articles, and group websites all of which are available to the public, including those individual who might be identified in one of the articles. No information will be collected directly from individuals and thus there is no opportunity for individuals to decline to provide information for use in GVIM.

6.3 Do individuals have the right to consent to particular uses of the information? If so, how does the individual exercise the right?

No. The individual does not have the right to consent to the use of information in GVIM. GVIM will only use information available to the general public. There is no direct interaction with the individuals and so no opportunity to consent to particular uses of the information.

6.4 Privacy Impact Analysis: Describe how notice is provided to individuals, and how the risks associated with individuals being unaware of the collection are mitigated.

The risk to individuals would be that they would be unaware that the information had been collected. This PIA is being published to mitigate that risk. This PIA provides general notice of the type of information that will be used in GVIM and the limited purpose for that use. Notice is not provided to the individual because there is no direct interaction between DHS and the individual in the development of GVIM or GVIM's use of information. Information is collected from open, published, publicly available sources. The use of the information is for research and development purposes only and will not be used to make operational decisions. No actions or decisions will be made that affect the individuals referenced in the PII used by GVIM. If I&A chooses to implement GVIM in an operational setting, I&A will develop a new PIA covering that separate operational use.



Section 7.0 Access, Redress and Correction

The following questions are directed at an individual's ability to ensure the accuracy of the information collected about them.

7.1 What are the procedures that allow individuals to gain access to their information?

Individuals can access the source data from the same publicly available sources used in the GVIM project. Individuals may not gain access to GVIM directly.

7.2 What are the procedures for correcting inaccurate or erroneous information?

Subject matter experts will review the information housed in the GVIM framework for accuracy by checking the information against multiple sources and relying on their own established expertise in the area. Individuals could presumably correct their own information by approaching the publisher of a given report and requesting a correction or retraction.

7.3 How are individuals notified of the procedures for correcting their information?

Individuals would need to contact the original publisher of the data to find out the procedures for correcting any erroneous information.

7.4 If no formal redress is provided, what alternatives are available to the individual?

The individual may contact the publisher and request a correction or retraction of any erroneous information.

7.5 Privacy Impact Analysis: Please discuss the privacy risks associated with the redress available to individuals and how those risks are mitigated.

The risk posed by GVIM to an individual's privacy associated with redress is mitigated by the fact that the information will not be used to make determinations about individuals and individuals may approach the original publisher to pursue redress if any information is inaccurate. As the information will be used by GVIM only for R&D and will not be used in any operational capacity, no direct actions will be taken regarding the individual based on information used through GVIM.



Section 8.0 Technical Access and Security

The following questions are intended to describe technical safeguards and security measures.

8.1 What procedures are in place to determine which users may access the system and are they documented?

The information is currently stored on a secure government network with controlled access, password protection, and a firewalled system.

8.2 Will Department contractors have access to the system?

Yes. Federal contractors researching and developing GVIM will have access to GVIM but only to the extent necessary to perform their specific work.

8.3 Describe what privacy training is provided to users either generally or specifically relevant to the program or system?

Basic privacy training was provided to all personnel with access to the information related to GVIM. This training outlines the requirements of the Privacy Act and the responsibility of all federal employees and contractors to safeguard PII.

8.4 Has Certification & Accreditation been completed for the system or systems supporting the program?

S&T is currently working with I&A regarding the C&A process.

8.5 What auditing measures and technical safeguards are in place to prevent misuse of data?

The data is housed in a secure government network with controlled access, password protection, and a firewalled system. Any suspected unauthorized access will be immediately reported as a privacy and/or security incident.

8.6 Privacy Impact Analysis: Given the sensitivity and scope of the information collected, as well as any information sharing conducted on the system, what privacy risks were identified and how do the security controls mitigate them?

Potential privacy risks have been mitigated by limiting access to and use of PII used to develop GVIM. Except in limited situations, only the research team collecting the detailed data will use any PII



that might be found in the data about groups, and then only to support the models of group behavior. Data associated with GVIM will be protected with passwords, controlled & limited access, a firewall, and in some cases, encryption.

Section 9.0 Technology

The following questions are directed at critically analyzing the selection process for any technologies utilized by the system, including system hardware, RFID, biometrics and other technology.

9.1 What type of project is the program or system?

GVIM is a research and development project designed to test the effectiveness of using social science to model and/or simulate emerging violent group behavior. GVIM will not be used to support operational decisions.

9.2 What stage of development is the system in and what project development lifecycle was used?

The system currently consists of separate modules maintained at separate National Labs. The modules are currently being integrated into a single system that will be transferred and housed on a single computer. The Transformation Lifecycle is being used and we believe the system is currently at TRL 4.

9.3 Does the project employ technology which may raise privacy concerns? If so please discuss their implementation.

The research team will use a key word search tool that technically could be used to search for a person's name. However, the research team will not use GVIM to direct searches for information about individuals and will only search for information about groups.

Approval Signature

Original signed and on file with the DHS Privacy Office

John W. Kropf
Acting Chief Privacy Officer
Department of Homeland Security