Lisa J. Sotto, Chair
DHS Data Privacy and Integrity Advisory Committee

September 29, 2014

The Honorable Jeh Charles Johnson
Secretary of the U.S. Department of Homeland Security
U.S. Department of Homeland Security
Washington, DC 20528

Ms. Karen L. Neuman
Chief Privacy Officer
U.S. Department of Homeland Security
Washington, DC 20528

Re: DHS Data Privacy and Integrity Advisory Committee: Privacy Recommendations Regarding Auditing and Oversight of the DHS Data Framework (Report 2014-02)

Dear Secretary Johnson and Ms. Neuman:

It is my pleasure to convey to you the enclosed report that sets forth recommendations for DHS to consider when conducting oversight and audits of the DHS Data Framework. The Committee appreciates the Department's consideration of our previous recommendations (DPIAC Report 2011-01 on Policy and Technology Recommendations for a Federated Information-Sharing System) and understands that a disassociated system would not support DHS operational needs. We also appreciate that the Department is optimizing the DHS enterprise through access controls, auditing, and analytical precision, and by controlling and safeguarding DHS information while supporting DHS operational needs. In the report enclosed herewith, we provide twenty specific recommendations, including the need for Secretary-level sponsorship and regular monitoring by authorized Department executives, such as the DHS Privacy Officer, to ensure robust oversight.

This report is the result of an extensive effort by Committee members working closely with DHS components to research the topic of the report. We are grateful for the Department's cooperation in providing programmatic justifications for the need for big data analytics and processing, and for making officials with direct knowledge and expertise on the matter available to us.

We hope you will agree that implementing these recommendations in connection with the Department's Data Framework will enhance the protection of personal information while maintaining the effectiveness of the Department's mission.

Please do not hesitate to contact me if you have any questions regarding these recommendations.

Sincerely,

Lisa J. Sotto

Attachment:

Report 2014-02 Privacy Recommendations Regarding Auditing and Oversight of the DHS Data Framework

cc: Members of the DHS Data Privacy and Integrity Advisory Committee

Report 2014-02 of the Data Privacy and Integrity Advisory Committee on Privacy Recommendations regarding Auditing and Oversight of the DHS Data Framework

As Approved in Public Session September 22, 2014

Task

On January 27, 2014, the DHS Data Privacy and Integrity Advisory Committee (DPIAC) received a request from DHS Chief Privacy Officer Karen Neuman to provide comment on the Department's technical and policy approach to Big Data. In particular, the Technology Subcommittee was asked to address auditing and oversight issues in Departmental Big Data projects as follows:

> *In developing audit capabilities in DHS Big Data projects, what specific activities should we seek to audit and how can we best build those requirements into the technology? What policies are needed to support the technology? Once the audit logs are developed, how do we use them in a meaningful way to ensure robust oversight? For example, should the audit logs contain the responses to queries or not? What process should be in place for approving new or updated access controls? What mechanisms should be in place to ensure that these controls are not circumvented? Similarly, what mechanisms should be in place to ensure access controls are not changed without appropriate oversight? What mechanisms should be in place to identify anomalies in the use of the system by individuals?*

DPIAC Recommendations

The DPIAC makes the following recommendations:

<u>With Respect to Auditing Activities:</u>

1. DHS should define and adopt key performance indicators (KPIs, i.e., system metrics) for both system use and system quality, with Secretary-level sponsorship and regular monitoring by authorized agency executives, in coordination with the Oversight Offices [e.g. the Privacy Office (PRIV), the Office for Civil Rights and Civil Liberties (CRCL), and the Office of General Counsel (OGC)]. These KPIs should include measures to assess precision, sensitivity, and specificity for systems that tag attributes, link records, or provide search results.
2. DHS should ensure that audit log content contains enough information to accurately reconstruct previous operational scenarios and to support continuous system improvement.
3. To support investigations regarding specific decisions, audit logs should have the capability to "replay" identical query sessions in their entirety, including both queries and their results. Thus, queries should be retained in audit logs, even if they contain personally identifiable information (PII). However, given the data volume and volatility of the underlying query results, implementations that satisfy this part of the

recommendation may vary. Query results may be retained in the logs, referenced within the logs, or provided by some other mechanism as long as identical query sessions can be replayed in their entirety.

4. DHS should define reasonable policies to provide timely redress in the case of precision errors.
5. DHS should ensure that audit content supports adopted KPIs, where feasible.
6. DHS should adopt a lifecycle management approach to its audit log program.
7. DHS should adopt feasible, enforceable policies regarding audit logs, with Secretary-level sponsorship and agency-wide management support, including a policy of storing Big Data logs in a form that preserves the originally collected data and utilizes centralized log collection to create immutable logs and to prevent log tampering.
8. DHS should document and implement processes to support audit log management and review, including workforce training.
9. On a periodic basis, at minimum annually, DHS should engage in a formal, documented evaluation and effectiveness review of the Big Data system(s), adopted KPIs, and the audit log program.
10. DHS should invest in and use automated mechanisms such as audit log correlation, aggregation, or consolidation tools that complement a skilled workforce to inspect selected records, or linked records from multiple systems, and perform audit log analysis.
11. DHS should commit resources for the cleaning and ingestion of log files into a centralized, tool-based infrastructure, using standardized categories relative to the KPIs, different types of actions, and standard response processes for each type of action.
12. DHS should implement controls to protect the confidentiality and integrity of audit logs.

<u>With Respect to Oversight Activities:</u>

1. DHS should implement common access control safeguards for all Big Data projects/systems:
    a. Big Data systems should internally maintain the identity of all active users and be able to link actions to specific users.
    b. Ensure that all user credentials belong to currently authorized users.
    c. Inactive credentials should be disabled after a specific period of time.
    d. Require users to authenticate their claimed identities on information technology systems.
    e. Limit the number of log-on attempts and enforce account lockout conditions.
2. DHS may also want to consider enabling access restrictions, to particular system resources, or developing audit log reporting, based upon physical or logical location, time-of-day and day-of-the-week/month, and device used (e.g., mobile devices) where appropriate for specific agencies and roles.
3. DHS should define access control procedures in a manner consistent with specifications made in the DHS Data Framework, by the Common Vetting Task Force (CVTF), or the DHS Information Sharing and Safeguarding Governance Board (ISSGB) (or successors to these task forces and governing boards that may be created that include executive level sponsorship and oversight from the Privacy Office, Office for Civil Rights and Civil Liberties, and Office of General Counsel).

4. DHS should implement specific safeguards for the administration of all access control policies, subject attributes, and object attributes:
    a. All access control related actions should be monitored and logged on secure WORM (Write Once Read Many) devices.
        i. The log entries shall provide a non-repudiable record of which specific ID performed, or attempted to perform, each specific action.
        ii. These logs shall be maintained securely (e.g., encrypted, stored offsite, etc.) and protected from accidental or purposeful destruction or disabling.
        iii. This logging capability shall be made highly available and resilient against outages or attack.
    b. Access control administration credentials may never be shared under any circumstances.
    c. All access control changes, particularly those broadening existing access privileges, adding new access privileges, or adding/deleting/updating new attributes or roles, must be privileged capabilities, attributable to a specific individual, securely logged (see 4a above), automatically detectable, and fully reversible.
5. DHS should create and document its change control processes, including processes for making and approving an initial access request, system requirements, testing results, and approvals.
6. Reviews should be regularly conducted to make sure access control additions, changes and deletions have been managed according to established policy.
7. DHS should establish a baseline for normal access control behavior and monitor for anomalies against that baseline.
8. DHS should invest in and use automated mechanisms to monitor employee access to Big Data systems and actions within those systems by enabling correlation across the variety of information sources.

The DPIAC recognizes that many of these recommendations may require the acquisition of additional staff, tools, or services in order to implement. As such, the DPIAC also recommends that funding be provided to implement these recommendations.

Overall Findings

To respond to this tasking, the Technology Subcommittee requested briefings and/or demonstrations of the Neptune pilot, Common Entity Index Prototype (CEI Prototype), and Cerberus pilot. The Department provided materials and access to a broad cross-functional selection of personnel from agencies that have been involved in the projects.

The demonstrations and briefings were informative and on-point to provide the Subcommittee necessary background and usage information. We appreciate the time and effort of the Department's personnel in working with the Subcommittee on this tasking request.

To respond to the tasking, the Technology Subcommittee addressed the auditing and oversight issues separately.

Findings Related to Audit:

> *In developing audit capabilities in DHS Big Data projects, what specific activities should we seek to audit and how can we best build those requirements into the technology? What policies are needed to support the technology? Once the audit logs are developed, how do we use them in a meaningful way to ensure robust oversight? For example, should the audit logs contain the responses to queries or not?*

The DPIAC supports and encourages DHS to consider implementing the "key practices" identified by the National Institute of Standards and Technology (NIST) as necessary to meet the "challenges" associated with log management: (1) prioritize log management appropriately throughout the organization; (2) establish policies and procedures for log management; (3) create and maintain a secure log management infrastructure; and (4) provide adequate support for all staff with log management responsibilities.[1] The DPIAC's recommendations are organized around the key areas of content, policy and process, and tools.

The DPIAC finds that:

- Audit logs serve many functions including, but not limited to, troubleshooting problems, optimizing systems or networks, recording user actions, and providing useful data about potentially inappropriate activities. In this document, the terms "audit logs" and "log data" mean event information such as successful or failed search queries (and potentially, feedback on their results), data access attempts, account changes, and use of privileges.
- Audit logs should be viewed as only one component of a robust audit and oversight program. While the scope of this task focused on the effective use of audit logs and log data, the DPIAC encourages DHS to consider supporting a more broadly defined auditing and investigations function to govern its Big Data projects.
- Historically, audit logs have been used by government agencies as a validation tool to verify compliance with regulatory or statutory requirements; as a method to conduct investigations on wrongful system access; or as an incident tracking system. The advent of large data set aggregation has expanded the role of audit logs, making them the bedrock of reporting and metrics gathering, as well as developing predictive analytics to implement countermeasures against wrongful data use.
- Defining and maintaining KPIs can enable data-driven, operationalized audit activities (in contrast to episodic or incident-level audits alone) that provide oversight and governance of both system use and system quality.

Content

Audit content should support proper operational audit of the system's use and quality. Operationalized auditing for both system use and system quality is typically achieved through the development, capture, and review of KPIs. While they may be sourced from both audit log and other data, using KPIs forces the distillation of voluminous audit logs into stable,

[1] National Institute of Standards and Technology (NIST), Special Publication 800-92, *Guide to Computer Security Log Management*, Sept. 2006, pgs. 2-10, 11, available at http://csrc.nist.gov/publications/nistpubs/800-92/SP800-92.pdf ("NIST Guide to Log Management"). Note that NIST includes application-level logs, as those contemplated by DHS Big Data projects, within the scope of its guidance.

manageable metrics that can be monitored over time by DHS, at all levels. Typical usage scenarios might include analyzing operational efficiency; investigating suspicious activity (*e.g.*, event characterization and severity); and performing regular effectiveness reviews. Typical quality scenarios might include assessment of system uptime/response time; delivery of incorrect information; and missed delivery of correct information. Audit log content should contain enough information to accurately reconstruct previous operational scenarios as well as support continuous system improvement.

Proper system-use audit logs must contain basic system access authority/approval as well as search queries. Audit logs must contain the basics of who accessed the system, when and from where they accessed the system, and in what role they acted while accessing the system. Access events should include cases of automated machine-to-machine access. For ease of review, all access-related logged events should be closely correlated to current access controls.

Both queries and their responses may contain personally identifiable information (PII). To protect privacy and enhance data security, queries could be scrubbed of PII and responses not be retained in audit logs. However, such a policy would not support meaningful oversight or improved system quality. For example, if search terms containing PII are not retained, inappropriate queries may go undetected. In contrast, fully retained queries could be “replayed” to support investigations regarding specific decisions. Thus, the DPIAC recommends that these competing issues be balanced by retaining the queries, even if they contain PII, but query results need not be retained in audit logs as long as identical results can be reproduced for a given query.

System quality is vitally important to DHS systems since they directly bring the force of government into people’s lives. Thus, audit logs should support continuous improvement of systems that tag attributes, link records, and provide search results. The DPIAC recommends that the KPIs include measures to continually assess precision (e.g., incorrect CEI matches, Neptune tags, and Cerberus search results), sensitivity (e.g., missed CEI matches, Neptune tags, and Cerberus search results), and specificity (i.e., audit false alarms). Such KPIs are inevitably in tension, and technology alone cannot resolve the inherent conflicts among them. Specific policy trade-offs must be made.

For example, highly sensitive profile matching systems have lower precision because, though they miss fewer matches, they more often match the wrong people, creating greater privacy impact. By contrast, profile-matching systems with higher precision are less sensitive because profile matches are only made if they meet a more stringent standard. An example of a high precision, low sensitivity system would be one that matches profiles by exact name, birth date, and current address. High precision systems have lower privacy impact but are operationally less effective.

To minimize the impact of such data errors, the DPIAC recommends that DHS define reasonable policies to provide timely redress in the case of precision errors.[2] To reduce system errors over time, DHS may consider allowing users to flag errors for capture. For example, a field agent

[2] See DPIAC Report 2010-01, *The Elements of Effective Redress Programs*, March 25, 2010, available at http://www.dhs.gov/xlibrary/assets/privacy/privacy_dpiac_report2010_01.pdf). Similarly, the Fair Credit Reporting Act of 1970 (FCRA) mandates handling cases of mistaken identity in its “adverse information” provisions.

notices that the system missed that two profiles are really the same person (a sensitivity error) or that a person is a mistaken composite of two different people (a precision error). By capturing user feedback, system analysts could assess such errors and, where verified, resolve and incorporate them into a growing set of truth data used for ongoing system improvement.

Policies & Processes: Audit Log Lifecycle Management

The DPIAC recommends that DHS adopt a lifecycle management approach to audit log management. Such an approach can provide effective governance and oversight for Big Data projects, while striking a reasonable balance among available resources, the need for continuous improvement in the underlying systems and audit capabilities, and the inevitable, onslaught of log data (see Figure 1).[3]

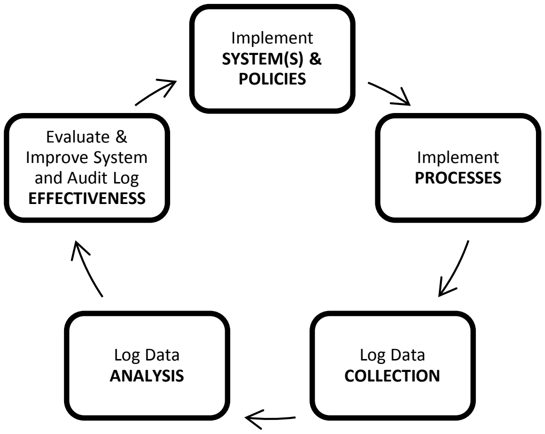
Implement SYSTEM(S) & POLICIES
Implement PROCESSES
Log Data COLLECTION
Log Data ANALYSIS
Evaluate & Improve System and Audit Log EFFECTIVENESS

Figure 1. Audit Log Management Lifecycle

Initially, DHS should develop policies regarding audit logs that address, at a minimum, content management, including data collection and storage; appropriate use, including analytics use cases; roles and responsibilities; access and authorities; protective measures, including the preservation of log data in its originally collected form; event response, including prioritization and investigations; and data retention. Policies must be feasible and enforceable, with clear Secretary-level sponsorship and agency-wide management support.

Prior to moving into an operational mode, DHS should document and implement supporting processes that, at a minimum, address policy enforcement; infrastructure support; and analytics use case development and maintenance. Analytics use cases should incorporate both system use and system quality perspectives and include, at the least, automated alerts and reporting for system usage and anomaly detection, defined analyst review scenarios, and authorization for ad hoc review. DHS should train its workforce regarding proper use of log data and analytics.

The DPIAC also recommends that DHS define and collect a set of KPIs to measure the effectiveness of the audit log program and the Big Data system(s) themselves, including data quality, system usage, and the investigation and outcome of prioritized audit events. KPIs should be designed to enable data-driven governance, oversight, and continuous improvement, and should be selected with clear Secretary-level sponsorship and management support. The DPIAC recommends that the Secretary and other authorized agency executives regularly monitor and review these important measures of DHS Big Data projects. Through management oversight and assigned accountability, DHS should also ensure that the core operational activities to collect and analyze audit log data are performed consistently with the policies.

[3] The approach recommended here is consistent with the "Audit and Accountability" Controls Family in NIST Special Publication 800-53, Rev. 4, *Security and Privacy Controls for Federal Information Systems and Organizations*, and the NIST Cybersecurity Framework, developed pursuant to Executive Order 13636.

Finally, on a periodic basis, at minimum annually, DHS should engage in a formal, documented evaluation and effectiveness review of the Big Data system(s) and the audit log program, including assessment of the KPIs; user feedback regarding data quality and usability; event response and investigations activities; and any identified policy gaps or other lessons learned throughout the review period. Findings should be used to make appropriate updates and define additional or modified requirements, all in a manner that supports continuous improvement.

Tools

The DPIAC recommends that DHS invest in and use automated mechanisms such as audit log correlation, aggregation, or consolidation tools that complement a skilled workforce to inspect records and perform audit log analysis. Automated mechanisms alone are not a replacement for skilled personnel for analyzing log data, and should not be treated as such. Audit log tools should assist with the manual detection of unusual activities and provide quick anomaly identification. Selected automated mechanisms may range from common free log management utilities to vendor-supported log management tools at the system (application) level, operating system level, or component level.

The DPIAC recommends that DHS develop standard categories of audit log records relative to the KPIs, different types of actions, and standard response processes for each type of action. Audit logs should include relevant information such as time stamps, source and destination addresses, user/process identifiers, event descriptions, success/fail indications, filenames involved, and access control or flow control rules invoked. There is typically little consistent format (e.g., XML, json, syslog) or schema (field naming, types) within log data. Therefore, DHS should commit resources for the cleaning and ingestion of log files into a centralized, tool-based infrastructure to create immutable logs and prevent log tampering, utilizing a standardized data representation that ensures that the data is normalized and categorized consistently.

Audit log records should be retained until they are no longer needed for administrative, legal, audit, or other operational purposes. This includes, for example, retention and availability of audit log records relative to Freedom of Information Act (FOIA) requests, subpoenas, and law enforcement actions. Lastly, the DPIAC recommends that DHS implement controls to protect the confidentiality and integrity of the logged information.

Findings Related to Oversight:

> *What process should be in place for approving new or updated access controls? What mechanisms should be in place to ensure that these controls are not circumvented? Similarly, what mechanisms should be in place to ensure access controls are not changed without appropriate oversight? What mechanisms should be in place to identify anomalies in the use of the system by individuals?*

The DPIAC finds that:

- Access control systems are put in place to restrict access to systems and data contained within systems. Such systems limit access to data to authorized users with legitimate needs to use that data and who have received appropriate data use and protection training.
- Access control systems are critically important to Big Data projects to promote data sharing amongst disparate DHS components.
- Change control processes must be a part of any access control system to ensure that changes are not made to Big Data systems that would allow more (or less) access to data in those systems than the DHS Data Framework, Common Vetting Task Force (CVTF), DHS Information Sharing and Safeguarding Governance Board (ISSGB), and current oversight bodies allow (or successors to these task forces and governing boards that may be created that include executive level sponsorship and oversight from the Privacy Office, Office for Civil Rights and Civil Liberties, and Office of General Counsel).
- DHS would be best served by implementing anomaly detection practices to identify unexpected access control results that do not conform to rules implemented by the DHS Data Framework, Common Vetting Task Force (CVTF), or the DHS Information Sharing and Safeguarding Governance Board (ISSGB) (or rules implemented by any successors to these task forces and governing boards that may be created that include executive level sponsorship and oversight from the Privacy Office, Office for Civil Rights and Civil Liberties, and Office of General Counsel).

Access Control Systems

Access control systems are put in place to restrict access to systems and data contained within systems. In general, such systems limit access to data to authorized users with legitimate needs to use that data and who have received appropriate data use and protection training. As requested, the Subcommittee also considered what oversight activities might be necessary in granting, maintaining, reviewing, and identifying anomalies in access control systems.

Given the sensitivity of the data and the level of personal information DHS Big Data projects will consume, a more centralized approach may be appropriate to determine access controls. However, it was noted in the DHS *Privacy Impact Assessment for the DHS Data Framework* that, "DHS will change access control from the existing Role Based Access Control (RBAC) approach to one that includes…enforcement of who (User Attributes) is allowed access to individual data elements (Data Tags) for particular purposes (Context = Purpose + Function)."[4] This type of attribute based access control (ABAC) has been called out as a "recommended access control model for promoting information sharing between diverse and disparate organizations."[5] ABAC is a very precise method of access control where "subject requests to perform operations are granted or denied based on assigned attributes of the subject, assigned

[4] Department of Homeland Security, *Privacy Impact Assessment for the DHS Data Framework*, November 6, 2013, available at: http://www.dhs.gov/sites/default/files/publications/privacy-pia-dhs-wide-dhsdataframework-11062013.pdf.

[5] NIST Special Publication 800-162, *Guide to Attribute Based Access Control (ABAC) Definition and Considerations*, January 2014, available at: http://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.800-162.pdf.

attributes of the object, environment conditions, and a set of policies that are specified in terms of those attributes and conditions."[6] Processes (and related documentation describing such processes) that must be put into place to ensure appropriate access control under this scheme include:

- Proper credential issuance
- Credential validation and protection
- Authoritative subject and object attributes (user attributes and data tags)
- A common attribute taxonomy
- Rule management for access control decisions
- Audit mechanisms that track access of objects to specific subjects that are linked to specific users.[7]

These processes should be regularly reviewed for efficacy and continued appropriateness by the Oversight Offices (e.g. the Privacy Office, the Office for Civil Rights and Civil Liberties, and the Office of General Counsel).

Maintenance of access controls is critical. Changes in the employee population (new employees, changing roles, and departures), as well as changes in reporting structures, new processes, new systems and new business requirements, new data tagging taxonomies can all create instances of non-compliance and risks to information if not properly managed.

Change Control

The DPIAC recommends that DHS implement a robust change control process related to departmental Big Data projects. Change control processes ensure that changes to information technology systems are made in a controlled manner that does not otherwise interrupt the proper and expected operation of the underlying system. Appropriate change control procedures and sign-offs should be included before any type of system changes can be put into production. Changes that must be monitored and tracked in a Big Data access control model may include, but are not limited to:

- Changes to user access levels
- Addition or deletion of source systems included in any Big Data repository
- Changes to user attributes, automated query policies, data tagging, etc.

Additionally, the DHS would be best served by adopting a "Three Lines of Defense" model for access control oversight.[8] This model is used in many situations to prevent risk to an organization. In most situations, the first line of defense is trained business staff and operational activities, the second line of defense is oversight functions (policy and process), and the third line of defense is audit and review.

[6] *Id.*

[7] *Id.*

[8] Institute of Internal Auditors, *The Three Lines of Defense in Effective Risk Management and Control*, January 2013, available at: https://na.theiia.org/training/templates/Pages/The-Three-Lines-of-Defense-in-Effective-Risk-Management-and-Control.aspx.

In this model, operational departments are the first line of defense. This line is responsible for implementing and maintaining the access controls. With the size and complexity of DHS, we recommend that a cross-functional group, similar to the Common Vetting Task Force, be considered as the first line representative. The Oversight Offices (e.g. the Privacy Office, the Office for Civil Rights and Civil Liberties, and the Office of General Counsel) could be considered the second line of defense, which provides oversight, input, and training for the first line regarding regulatory and policy requirements. This second line also provides monitoring and testing to see that the access controls, processes, and procedures for compliance are in place and operating effectively. The third line of defense is the DHS Office of the Inspector General, which provides formal audits of the activities of the first two lines of defense. This supports proper segregation of duties and maintains clear accountabilities.[9]

Mechanisms to Identify Anomalies

Access control systems depend on predictable results. Results that are not intended (either through design or if implemented in the absence of effective change control practices) can result in unauthorized access and disclosure, fraud, and sabotage or denial of service.[10] In general, it is difficult to determine *a priori* whether a particular attempted unauthorized action is due to error, ignorance, or malice. Implementing anomaly detection practices to identify unexpected access control results that do not conform to established rules must be considered a best practice of any access control method.

Next Steps

The DPIAC appreciates the opportunity to provide the Department with these recommendations regarding audit and oversight of DHS Big Data projects. We encourage DHS to implement these enhanced audit and oversight measures as the Framework moves further into production.

[9] Throughout this paper, the Subcommittee is considering access and logging of the Big Data system should be handled and approved separately and apart from access to the source systems.

[10] For the purposes of this paper, we are primarily considering anomaly detection related to insider threats (threats introduced by users with some sort of legitimate access to DHS Big Data systems). The Subcommittee highly recommends a reading of The Carnegie Mellon University Software Engineering Institute *Common Sense Guide to Mitigating Insider Threats* 4th ed. (Dec. 2012). This guide provides additional information that defines these threats and provides mitigation solutions. See http://resources.sei.cmu.edu/asset_files/TechnicalReport/2012_005_001_34033.pdf

Privacy Office
U.S. Department of Homeland Security
Washington, DC 20528-0655

Homeland
Security

January 27, 2014

Ms. Lisa Sotto
Chair, DHS Data Privacy and Integrity Advisory Committee (DPIAC)
c/o Hunton & Williams LLP
200 Park Avenue
New York, NY 10166

Re: DPIAC Guidance on Transparency and Oversight of the Department's Use of Big Data

Dear Lisa,

The Department of Homeland Security (DHS) continues to evaluate its technical and policy approach to Big Data, including data governance and information sharing. The DPIAC's 2011 recommendations on the subject continue to inform our decisions as we strive to support the Department's operational needs while protecting individual privacy. As discussed at the September 12, 2013 meeting, the Department's understanding of its data practices and operational environment (both from a mission perspective and technical perspective) is evolving.

As a result, we were unable to implement a key recommendation from your paper, *DPIAC Report 2011-01*. Specifically, the Committee recommended that in order to mitigate privacy risks, DHS minimize central storage of information when implementing its Big Data system – so long as there is little or no reduction in the effectiveness of the mission. Over the last two years, DHS determined that a system with minimal central data storage is not operationally feasible and, significantly, will not allow for effective privacy controls. Therefore, DHS has created Neptune, a central repository on the unclassified network; Cerberus, a central repository on the classified network; and the Common Entity Index Prototype (CEI Prototype), an identity resolution capability. These are all part of a series of pilot programs that are described in detail in four PIAs and one SORN.

The Privacy Office now seeks additional guidance from the DPIAC on: (1) whether and how to increase transparency; and (2) what additional oversight mechanisms should be implemented as the pilots move to operational programs.

In order that the Department might benefit from the substantial expertise and knowledge of the Committee members, I ask that the DPIAC provide written guidance about privacy best practices for notice and transparency related to our use of Big Data, including information sharing with other agencies, and the use of audit mechanisms in the oversight process. Specifically, I ask that the Committee consider and address the following:

Notice/Transparency

> In addition to the already published PIAs and SORN, and future updates to those documents, what should DHS consider doing to expand and improve notice to the public? For example, should Privacy Act notices provided at the point of collection be revised to address Big Data in some way, including repurposing of data in source systems, and/or are there means of notice that could be provided other than that specifically required by the Privacy Act and the e-Government Act?

Auditing/Oversight

> In developing audit capabilities in DHS Big Data projects, what specific activities should we seek to audit and how can we best build those requirements into the technology? What policies are needed to support the technology? Once the audit logs are developed, how do we use them in a meaningful way to ensure robust oversight? For example, should the audit logs contain the responses to queries or not? What process should be in place for approving new or updated access controls? What mechanisms should be in place to ensure that these controls are not circumvented? Similarly, what mechanisms should be in place to ensure access controls are not changed without appropriate oversight? What mechanisms should be in place to identify anomalies in the use of the system by individuals?

I ask that the Policy Subcommittee address transparency and that the Technology Subcommittee address oversight by engaging in fact-finding to support a public report and recommendations from the Committee addressing these important issues. If my office can provide any assistance to you as the Committee undertakes this tasking, please do not hesitate to let Shannon Ballard or me know.

Very truly yours,

Karen L. Neuman
Chief Privacy Officer