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: Guidance on Transparency and Notice in the Department of Homeland Security Data Framework (Report 2014-01)

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 regarding notice and transparency related to use of the DHS Data Framework, including information sharing with other agencies. 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 notes recent updates to relevant publicly posted Privacy Impact Assessments. In the report enclosed herewith, the Committee provides three specific recommendations, including increased efforts to provide greater transparency as to current and future uses of the relevant personal information.

This report is the result of an extensive effort by Committee members working closely with DHS components to research the relevant topic. 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 2014-01 Guidance on Transparency and Notice in the DHS Data Framework

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

Report 2014-01 of the DHS Data Privacy and Integrity Advisory Committee on Guidance on Transparency and Notice in the Department of Homeland Security Data Framework

As Approved in Public Session September 22, 2014

Background

The DPIAC has been asked to provide recommendations to the Department of Homeland Security (DHS) on considerations for expanding and improving notice to the public regarding the DHS's new Data Framework. The Data Framework is being developed to enable the more effective, efficient and controlled use of available departmental information across the DHS enterprise. The Framework is intended to alleviate mission limitations of the present system in which data systems are distributed across multiple components in DHS. It is also intended to have capabilities that will support advanced data architecture and improved governance processes, including processes to protect the privacy, civil rights and civil liberties of data subjects.

In 2011, the DPIAC provided privacy policy recommendations to DHS for a federated information-sharing system.[1] Specifically, the DPIAC recommended that in order to mitigate privacy risks, DHS minimize central storage of information when implementing such a 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.

As explained in the Privacy Impact Assessment (PIA), the Data Framework will enable a user to search an amalgamation of data extracted from multiple DHS systems for a specific purpose and view the information in an accessible format.[2] In addition to more efficient search and analysis capabilities across DHS data, the Framework will enable controlled information sharing in a manner that manages limits on search parameters and user access.

The Framework's elements for controlling data are: user attributes; data tags to label data types; origin and date of ingestion; context based on type of search and analysis and the purpose for use of the data; and dynamic access control policies based on user attributes, data tags and context. DHS is pilot testing three of these elements: data tags, context and dynamic access controls.

The Privacy Office now seeks additional guidance from the DPIAC about notice and transparency related to the use of the Data Framework, including information sharing with other agencies, and the use of audit mechanisms in the oversight process.

[1] Privacy Policy and Technology Recommendations for a Federated Information-Sharing System, Report No. 2011-01, (December 2011), available at www.dhs.gov/privacy-office-dhs-data-privacy-and-integrity-advisory-committee.
[2] Privacy Impact Assessment for the DHS Data Framework, DHS/ALL/PIA-046, November 6, 2013, available at www.dhs.gov/publication/dhsallpia-046-dhs-data-framework.

Specifically, the DPIAC Policy Subcommittee was asked to address the following aspects of Notice/Transparency:

- In addition to the already published PIAs and System of Records Notices (SORNs) and future updates to those documents, what should DHS consider doing to expand and improve notice to the public?

- Should the 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?

The DPIAC has examined notice and transparency in relation to the three DHS pilot programs and also the use of the Data Framework more generally.

I. Introduction

In order to get an overview of current practices regarding the pilots and the Framework and their relation to notice and transparency, the Policy Subcommittee received briefings from DHS staff on the three pilots designed to test aspects of the Framework. The briefings supplemented the information provided in the January 2014 public meeting of the DPIAC, and allowed for additional information and discussion on some of the security safeguards and auditing capabilities that are of a sensitive nature.

The three pilots tested controls implemented to mitigate security and privacy risks, as well as provided information on the benefits and risks associated with a federated search capability through the centralized Data Framework. The pilots are discussed below.

In addition to the briefings and the relevant SORNs and PIAs, the Policy Subcommittee also reviewed the current privacy notices from the forms used to collect personally identifiable information for the datasets involved in the pilots.

II. Big Data and Technology Enhancements

Harnessing the power of big data through enhanced computing techniques, searching across multiple and disparate databases, and maintaining a centralized data warehouse will enable DHS to provide more relevant and timely information to its customers – in a more secure and more privacy-respectful manner.

a. Summary of the Pilots

As described in the PIAs and SORNs, the Neptune Pilot, Cerberus Pilot and Common Entity Index (CEI) Prototype work together to streamline the searching of various individual systems in a holistic and more complete manner based on user attributes, data tags, contextual searches based on function and role-based access controls that are all auditable through immutable log history.[3] No longer will properly authorized officials have to manually search multiple databases and compare by hand the different or conflicting responses they receive.

DHS is testing this capability by ingesting data from three databases[4] into a new Neptune database where the data is tagged. The Neptune Pilot, residing in the Sensitive but Unclassified (SBU) domain, was designed to consolidate data in a repository and test the tagging of the data. The tagged data in the Neptune repository is used for the CEI Prototype and the Cerberus Pilot, but not for other purposes.

The Cerberus Pilot, residing in the Top Secret/Sensitive Compartmented Information (TS/SCI) domain, receives tagged data from Neptune. Cerberus was designed to test the ability to ensure that only users with certain attributes are able to access data based on defined purposes using the dynamic access control process. In addition to authorized purpose and use, Cerberus also tests the ability to perform simple and complex searches across different component datasets using different analytical tools.

The CEI Prototype, residing on the SBU domain, receives a subset of the tagged data from the Neptune Pilot and correlates data across component data sets. The Prototype was designed to test the utility of data tagging, specifically the ability to ensure that only users with certain attributes are able to access data based on defined purposes using dynamic access controls.

b. Benefits of the DHS Data Framework

The three pilot programs are designed to work together to optimize DHS search capabilities, improve the accuracy of analytics, safeguard the underlying data and accelerate the intelligence lifecycle. In addition, the technology employed to protect and secure the data also serves as Privacy Enhancing Technology, by enabling a more effective system of data access and use limits. The approach also provides greater transparency into searches, enhanced auditing and therefore improved accountability.

Specifically, the DHS Data Framework defines four elements for controlling data:
(1) User attributes identify characteristics about the user requesting access such as organization, clearance and training;
(2) Data tags label the data with the type of data involved, where the data originated and when it was ingested;
(3) Context combines what type of search and analysis can be conducted (function), with the purpose for which data can be used (authorized purpose); and

[3] DHS has published SORNs and PIAs on the Data Framework and three pilot tests: Neptune, Cerberus, and the CEI Prototype), all of which are discussed below in this paper. The documents are available at www.dhs.gov/privacy-documents-department-wide-programs.

[4] The databases are the Transportation Security Administration's Alien Flight School Program, the Customs and Border Protection's Electronic System for Travel Authorization, and Immigration and Customs Enforcement's Student Exchange and Visitor Information System.

(4) Dynamic access control policies evaluate user attributes, data tags and context to grant or deny access to DHS data in the repository based on legal authorities and appropriate policies of the Department.[5]

Through the DHS Data Framework, additional granularity is gained by controlling access based on the role and the specific function of persons conducting the search, the type of search, the purpose of the search and ultimately the authority the person has to search. Being able to harness this detail and audit it is a benefit not only to the security of the system, but also to the privacy of the data subjects.

III. Privacy Risks and Potential Notice Impacts of the DHS Data Framework

As previously detailed in a prior report, the DPIAC believes that there are risks that remain that are noteworthy and that could inform our recommendations regarding notice and transparency.[6]

Among these are six risks that are important to address:

- New Uses of the Data: As greater analytical capabilities are brought to bear on the data warehouse, a new segment of uses may be discovered that is outside the stated purposes of the current privacy notices.
- Privacy Notices: While the language of notices is written very broadly to accommodate many unanticipated uses, it is this uncertainty regarding potential new uses that could bring less transparency to a process and/or government functions that are designed to be the opposite.
- Format of the Notices: As uses of the data change over time or new, unanticipated uses become a reality, the format of initial and changed privacy notices becomes more important to provide transparency to the person whose data is being used. In many cases notices are only provided at the time of collection as opposed to when additional technology enhancements are introduced or new uses are made of the data.
- End User: As the population to which the privacy notices change or the status of a person changes under the notice provisions at the time, differences in the rights or protections afforded persons could change. As data is combined or introduced from multiple databases into one centralized repository, the legal status for each data point, the use or the laws protecting the data could be different. More specifically, the rights of U.S. and non-U.S. Persons may affect the initial notice and subsequent changes.
- Onward Uses: As it may be possible that users outside DHS gain access to this data through enhanced searching and a new use that was not previously considered comes about, it is important to ensure that this new user comports with the original DHS law enforcement and terrorism remits and with privacy notices. By being very specific and transparent about the uses, DHS may be able to avoid future problems.
- Redress: Finally, the DPIAC acknowledges that in a centralized data repository model the opportunity for incorrect data and processing to be present exists. Without

[5] *Supra*, note 1.

[6] *Ibid.*

providing some means of redress to correct the underlying system, the data repository will continue to propagate this incorrect data in future searches.

IV. DPIAC Recommendations

The main focus of the DPIAC in analyzing the three pilots and DHS Data Framework is to concentrate on: (1) additional methods by which privacy notices should be communicated, (2) how to communicate changes and (3) to whom these changes should be communicated.

The DPIAC is not as focused on the specific content of any one notice and also notes that its recommendations are extensible to other enhanced technologies and platforms separate from the three pilots mentioned herein.

The DPIAC recommends that DHS focus its efforts on (1) accessible notices, (2) living notices and (3) outreach activities to ensure transparency and trust in its privacy practices as new and/or enhanced technologies are brought online.

a. Ensure Notices Cover Multiple Affected Audiences

Key to the continued use and applicability of privacy notices that communicate trust, transparency and provide data necessary to convey the Fair Information Privacy Practices (FIPPs), the DPIAC recommends that its notices:

- Are tailored to or include applicable passages that apply to affected persons based on their legal immigration or citizenship status;
- Employ a variety of notice types that are accessible to persons with disabilities;
- Employ a variety of assistive technology media for those with limited literacy; and
- Present in different languages based on the populations being served and/or the geographic locality.

By providing broad-based notices that apply to multiple audiences, DHS will increase the likelihood that its notices can be understood by data subjects.

b. Create a Living Document

The most important recommendation that the DPIAC makes in regard to privacy notices is for DHS to provide ongoing updates and material changes through a living document that is consistently updated on its website. By providing a static website link on all privacy notices or other relevant materials, DHS can direct persons to its website for subsequent questions, a living FAQ section, more details and the practices described and versions of the notices in other languages and accessible formats. The website might also be used to provide updates to notices, pending changes to the notice on printed forms.

The website should give more specific information on how the data they have submitted are being used. This website would also provide services for affected persons, such as letting them sign up for ongoing updates, submit questions for DHS response in FAQs, review when policies change and receive further clarification of the programs that collect and use their data via links to PIAs and SORNs.

Furthermore, as in the current case of technology enhancements being piloted, DHS would have the ability to provide notice to the public in a readily available manner that, while it does not replace the Federal Register, may reach more persons and achieve the stated goals of better transparency. These efforts can be supplemented by some or all of the following: flyers and other notices at form distribution locations designed to lead affected persons to the website, contact with relevant industry groups and leaders that could direct persons to the website and the ability to "over-communicate" ahead of problems, leaks or other changes in technology. System objectives and metrics should also be provided, where appropriate.

c. Additional Transparency and Outreach

Finally, the DPIAC recommends that DHS reach out beyond the individual person and provide information on privacy notices, changes and other technology updates to other relevant organizations, such as the following:

- National and international associations that may be affected by notice changes (for example, shipping groups, railroad, airlines, truck drivers and other travel groups),
- Sector Coordinating Councils and other industry leader groups that are proficient at communicating changes and enhancements to their membership, and
- Privacy advocacy groups and professional privacy associations.

Although larger outreach to the community is usually reserved for the response to incidents, the DPIAC recommends that DHS adopt an over-communication strategy, including via news media, that is persistent (e.g., its website) and extensible to a wide audience prior to questions or problems being raised.

V. Conclusion

The DPIAC believes that while the existing privacy notices might meet the minimum legal requirements, we recommend that efforts be made to provide greater transparency as to the current and future uses of the collected personal information, including ensuring that the privacy notices are available in languages and formats that make them accessible to multiple audiences and making the notices and updates to them available on a website. We also recommend that DHS pursue continuous outreach activities to appropriate external groups to inform them of its data collection and use practices. We encourage DHS to implement these enhanced transparency measures as the framework moves further into production.

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