Report 2018-02 of the DHS Data Privacy and Integrity Advisory Committee on Office of Immigration Statistics (OIS) Data Dissemination Practices
As Discussed in Public Session
on December 10, 2018

Summary of Request and Context

The Department of Homeland Security (DHS) Data Privacy and Integrity Advisory Committee (DPIAC) established the Technology Subcommittee (the "Subcommittee") to explore the impact on the privacy of individuals where the conduct of DHS programs may require technology solutions or recommendations. The Subcommittee is further charged with offering recommendations to DHS to mitigate any privacy concerns raised by DHS programs.

I. Tasking

On September 15, 2017, the DHS Chief Privacy Officer issued a tasking requesting that the DPIAC provide guidance on how best to disseminate statistical data from the Office of Immigration Statistics (OIS) in order to strengthen the Department's ability to analyze the immigration enforcement and immigration benefits lifecycles, and provide real-time access to relevant immigration data needed to support operations, analysis, reporting, and strategic decision making. This integrated immigration data system will also support the Department's ability to disseminate statistics and to permit public access to statistical information that will inform key stakeholders and promote transparency.

Within this context, DPIAC was asked to provide a mix of high-level policy guidance on general data dissemination principles as well as technical guidance to:

1. Identify best practices for protecting data linked for statistical purposes, including "crosswalk" files containing identifiers, from both an information technology and policy perspective; and

2. Identify data disclosure methods, and whether is it advisable to consider variable controls for releases to different audiences/mediums. If such controls were utilized, what policy controls should be considered?

II. Additional Context

On May 7, 2018, DHS provided the following additional clarification:

- Regarding the first tasking question on best practices for data protection, the Subcommittee should discontinue its work on this question. OIS will rely on DPIAC Recommendations Report 2011-01. Privacy Policy and Technology Recommendations for a Federated Information-Sharing System regarding data protection best practices.
- Regarding the second tasking question on data disclosure methods, OIS is not requesting that the Subcommittee provide it with guidance on statistical methodology and proper

cell size regarding data disclosure. Instead, OIS is requesting the Subcommittee identify other policy concerns that should be considered regarding aggregate data disclosure given the very sensitive data underlying such aggregated disclosures.

III. Fact Finding

To complete this task, the Technology Subcommittee:

1. Received a briefing from Christa Jones, Senior Director, Privacy Policy and Oversight, and Michele Steinmetz, Program Manager, Office of Immigration Statistics, on the tasking and the integrated immigration data system project (November 30, 2017);

2. Reviewed the following documents provided by DHS in December 2017 regarding confidentiality and data access issues for federal agencies:

    a. Memorandum of Agreement Between the Department of Homeland Security Office of Policy, Office of Immigration Statistics and the Department of Justice Executive Office for Immigration Review Office of Planning, Analysis, and Statistics Regarding the Sharing of Information on Immigration Cases for Statistical Purposes (Undated draft received by the Subcommittee December 2017)
    b. Federal Committee on Statistical Methodology, Confidentiality and Data Access Committee, *Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology* (December 2015)
    c. Federal Committee on Statistical Methodology, Confidentiality and Data Access Committee, *Identifiability in Microdata Files* (July 2002)
    d. Confidentiality and Data Access Committee and Federal Committee on Statistical Methodology, *Confidentiality and Data Access Issues Among Federal Agencies* (November 2001)
    e. Federal Committee on Statistical Methodology, Confidentiality and Data Access Committee, *Paper presented at the FCSM Statistical Policy Seminar: Integrating Federal Statistical Information and Processes (*November 2000)
    f. Joint Statistical Meeting, *Panel on Disclosure Review Boards of Federal Agencies: Characteristics, Defining Qualities and Generalizability* (August 2000)
    g. Federal Committee on Statistical Methodology, Confidentiality and Data Access Committee, *Checklist on Disclosure Potential of Proposed Data Releases* (July 1999)
    h. U.S. Department of Commerce, *Statistical Policy Working Group Paper 2: Report on Statistical Disclosure and Disclosure-Avoidance Techniques* (May 1978); and

3. Received a briefing from Michael Hawes, Director of Student Privacy Policy Board, Department of Education, Federal Committee on Statistical Methodology (FCSM), Chair of the Confidentiality and Data Access Committee (CDAC) (January 2018).

IV. DPIAC Recommendation

The DPIAC offers the following response to the second tasking question regarding data disclosure methods.

- The Subcommittee recommends OIS conduct data re-identification testing to determine whether any publicly-provided statistical information could be positively re-identified to a unique individual. OIS should consider a conclusion that "re-identification is reasonably possible" to be the equivalent of an affirmative response on re-identification. Where re-identification is reasonably possible, the Subcommittee recommends that OIS consider additional techniques to further de-identify the data. Although the Subcommittee understands that it is difficult to ensure that data cannot be re-identified,[1] the Subcommittee believes OIS should pursue all reasonable de-identification techniques to prevent the re-identification of individuals' personal information.

  The Subcommittee has determined that it is important to ensure that all de-identification efforts are thoroughly conducted and tested, particularly where the data is of smaller cell size which increases the likelihood of re-identification. Best efforts should be made to utilize reasonable de-identification testing techniques that are recommended by the technology industry. Effort must also be made to ensure that inferences about data sets and individuals cannot be made across multiple sources of publicly-provided statistical information. OIS must make all reasonable efforts to ensure that the relevant data is no longer identifiable, and that inferences about the de-identified data could not be made, since possible uses of re-identified data could affect individuals' legal rights or cause them significant harm (e.g. in immigration enforcement cases).

  If data re-identification testing has not been considered, the Subcommittee recommends that a plan to conduct such exercises be developed and implemented on an ongoing and frequent basis. The plan should also consider the various data sharing and dissemination arrangements OIS may engage in and ensure data re-identification testing includes any expected datasets or variables. Following each such exercise, OIS should reconsider as appropriate whether additional de-identification techniques and safeguards should be implemented.

- The Subcommittee recommends that OIS consider that anonymization and aggregation actions alone may not be sufficient to protect privacy in data sharing agreements (e.g. the MOU provided to the Subcommittee) where other readily available variables, such as aggregated demographic data or geographic data, might be included in the data file. In certain situations, especially those where data sets are small in scope, it is possible that human-intelligence inferences could make correlations between the data variables leading to unintentional identification of individuals. For example, a comparison of the number of enforcement actions in a geographic area to a list of individuals staying in shelters in the same geographic area could allow for the deduction of the names of the individuals who are subject to enforcement action. OIS must use its best efforts to ensure that its data

[1] There have been documented cases of organizations releasing datasets that they thought they had anonymized well, only to have the dataset reidentified (e.g., the Netflix Prize data). See https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

cannot, when aggregated with other publicly available data, provide any user the ability to then access data that would de-anonymize the OIS dataset.

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

Homeland Security

September 15, 2017

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

Dear Lisa:

One of the Department of Homeland Security's (DHS) core responsibilities is to enforce the immigration laws and policies of the United States. The Department also has a duty to safeguard personally identifiable information (PII) in its possession, and ensure such protections are consistent with the DHS Fair Information Practice Principles (FIPPs). This obligation extends from DHS' mission to safeguard the American people, our homeland, and our values, which includes respect for privacy.

In recent months, the Department launched an Immigration Data Integration Initiative led by the Office of Immigration Statistics in the DHS Office of Policy and the Office of the Chief Information Officer in the DHS Management Directorate. This initiative seeks to establish an integrated immigration data system that will strengthen the Department's ability to analyze the immigration enforcement and immigration benefits lifecycles, and provide real-time access to relevant immigration data needed to support operations, analysis, reporting, and strategic decision-making. This system will also support the Department's ability to disseminate statistics and to permit public access to statistical information that will inform key stakeholders and promote transparency.

For a federal agency to disseminate statistical information, the undertaking requires certain commitments to government-wide data stewardship practices, including the acknowledgement and adherence to the principles and practices of statistical agencies, namely the protection of confidentiality and privacy.

Bearing this in mind, I am requesting that the Committee lend its expertise to the Department to provide advice to the Office of Immigration Statistics to identify best practices for compatible goals of safeguarding privacy and protecting confidentiality for immigration statistics data. Specifically, in the context of the integrated immigration data system, we are asking the Committee to:

- Identify best practices for protecting data linked for statistical purposes, including "crosswalk" files containing identifiers, from both an Information Technology and policy perspectives; and

- Identify data disclosure methods, and whether it is advisable to considerable variable controls for releases to different audiences/mediums. If such controls were utilized, what policy controls should be considered?

If my office can provide any assistance to you as the Committee undertakes this tasking, please do not hesitate to let Sandra Taylor, Christa Jones, or me know.

Very truly yours,

Philip S. Kaplan
Chief Privacy Officer