Report No. 2006-03

The Use of Commercial Data

This paper reflects the recommendations provided by the Data Privacy and Integrity Advisory Committee (Committee) to the Secretary and the Chief Privacy Officer of the Department of Homeland Security (DHS). The Committee's charter under the Federal Advisory Committee Act is to provide advice on programmatic, policy, operational, administrative, and technological issues relevant to DHS that affect individual privacy, data integrity and data interoperability and other privacy related issues.

I. Introduction

Our society is increasingly driven by and dependent on personal information collected by any number of organizations. It is no surprise to find that public agencies have an active interest in that same information. Personal information pertaining to our lives as citizens, customers, consumers, and community members is continuously collected, processed, used, and shared. Information about our finances, health, communications, behaviors and locations is increasingly integrated into comprehensive databases. These data sources operate across nearly every business and industry in our country, and increasingly the world.

Commercial data can reveal considerable information about individuals. For example, magazine subscriptions may provide insight into the political affiliations of the recipient. Durable goods purchases may reveal information about the individual's income. Location information may provide information about with whom an individual associates.

When these data elements are stored and processed by commercial entities, the resulting databases have a varying degree of data quality and integrity, because the initial purpose for the use of the data may not require a higher level of quality. The transparency of data systems and redress available to the individual from data holders vary as well. The use of these data for government purposes increases concerns, as the potential for harm to the individual increases with the unique powers of government to restrict people's rights, such as the ability to imprison. Due to this increased risk of harm, a higher level of data quality may be necessary.

Individuals have less knowledge about the use of this information since it is often not collected from them directly. Therefore, the use of commercial data by the government requires advance scrutiny. The Privacy Act supports this increased attention. The Act specifically sets forth the important principle that the government should "collect information to the greatest extent practicable directly from the subject individual when the information may result in adverse determinations about an individual's rights, benefits, and privileges under Federal programs."

One purpose of this paper is to encourage transparency, accountability and controls on the access and use of commercially acquired personal information for government decision-making. In this paper, we recommend policies and procedures for the DHS Privacy Office to guide, or limit, access to

commercial data by public agencies. In addition, we provide guidance on the use and management of such data.

II. Previous Committee Analysis

This document builds on previous work the Committee has done on the use of commercial data. In October 2005, the Committee published a report on the use of commercial data to reduce false positives in screening programs. The report is available on the DHS Privacy Office Web site at http://www.dhs.gov/interweb/assetlibrary/privacy_advcom_rpt_1streport.pdf. The Committee recommended that commercial data be used for screening programs only when:

- It is necessary to satisfy a defined purpose
- The minimization principle is used
- Data quality issues are analyzed and satisfactorily resolved
- Access to the data is tightly controlled
- The potential harm to the individual from a false positive misidentification is substantial
- Use for secondary purposes is tightly controlled
- Transfer to third parties is carefully managed
- Robust security measures are employed
- The data are retained only for the minimum necessary period of time
- Transparency and oversight are provided
- The restrictions of the Privacy Act are applied, regardless of whether an exemption may apply
- Simple and effective redress is provided
- Less invasive alternatives are exhausted

III. Recommendations:

This document builds on the Committee's prior work and specifically makes the following additional recommendations:

- The definition of Commercial Data should not exclude the following: (a) Publicly Available Data, data in the public domain that can be obtained or accessed from publicly accessible sources, both public and private; and (b) Public Record Data, data collected and maintained by a government entity for a public purpose and used outside of that public purpose.

- DHS should publish System of Records Notices (SORNs) for new or revised systems of records that use Commercial Data in a systematic manner or where there is substantial risk of harm.
- Apply PIAs to programs that use Commercial Data, where the Privacy Threshold Analysis (PTAs) shows Commercial Data is used systematically or where there is substantial risk of harm.
- Revise the PIA template and guidance documents to include a Commercial Data module and amend the analysis of completed PIAs where necessary.
- Have the DHS Privacy Office analyze the template contract language for Commercial Data vendor relationships, propose any necessary modifications, and review each relationship and contract.
- Make certain the DHS Privacy Office can effectively require the accurate and timely processing of PIAs, and mitigation of privacy risks noted therein.
- Make certain DHS commits sufficient resources to the DHS Privacy Office to (a) review the PIAs, (b) follow up to make certain privacy risks are mitigated, and (c.) ensure the PIA continues to be accurate as programs change.

IV. Definitions (and discussion thereof):

A. PII

Personally Identifiable Information ("PII") is defined by the DHS Privacy Office as, "information in a system, online collection, or technology: (i) that directly identifies an individual (e.g., name, address, social security number or other identifying code, telephone number, email address, etc.) or (ii) by which an agency intends to identify specific individuals in conjunction with other data elements, i.e., indirect identification.1

B. Commercial Data

The Committee defines Commercial Data as that PII accessed or obtained by a government agency from a non-governmental entity, including commercial and non-profit enterprises. Commercial Data may include some Publicly Available Data, and Public Record Data, when an individual would have a reasonable expectation of privacy as to the use of that information (described in more detail in the definitions of Publicly Available Data and Public Record Data) There are some exceptions, including situations where there is an existing legal obligation for the non-governmental enterprise to provide the data to the government agency, such as when employers must report payroll information, as long as such information is used solely for that purpose. The practice of government agencies transferring

[1] Privacy Impact Assessment Guidance, March 2006
http://www.dhs.gov/interweb/assetlibrary/privacy_pia_guidance_march_v5.pdf, P.10

commercial data between and among themselves does not change the characterization, as the important element is how the data originally became known to the government.

C. Publicly Available Data

Data in the public domain can be obtained or accessed from many public sources, making the task of differentiating between public and private data difficult, if not impossible. Also, given the ubiquity of the internet, and the ease with which individuals can post personal information about others, the meaning of publicly available information has changed dramatically over the past decade. For example, an individual could start a website providing photographs of individuals' who have been seen taking a certain book out of libraries. This website could allow other individuals to add to this list, and even encourage interested individuals to start visiting libraries to see who takes out the book. Once the information is posted to the internet, it could be considered Publicly Available Data. Also, when a government agency employee accesses data (either at a vendor, from a vendor, or directly (e.g. using an internet search engine)) it will in many cases be impossible for the employee to ascertain the original source of the data. Without being able to ascertain the original source, it is difficult to determine whether the individual would have a reasonable expectation of privacy on the use of this data.

Because governments may use publicly available data to individuals' detriment, and the difficulty in separating Publicly Available Data from other Commercial Data, it is appropriate to include such data in the definition of Commercial Data. The DHS Privacy Office's Privacy Impact Assessment (PIA) guidance currently recognizes the potential harm resulting from the government's use of this data, as it does not make an exception for Publicly Available Data.2

D. Public Record Data

Some information is intended to be held by the government. Examples of such public records are birth and death records, property records, motor vehicle and voter registrations, criminal records and civil case files. However, such information is provided to the government for a specific purpose (e.g. to register to vote). Use of this Public Record Data for purposes which the individual should not reasonably expect creates concern. The determination of whether an individual should have a reasonable expectation of privacy with respect to this data will depend upon the public's understanding of how the government has traditionally used the data. For example, governments often publish voter registration and some court records for all to use in any lawful way they choose. Whether there should be specific limitations on government's use of this data (e.g. using voter registration data to withhold benefits from individuals) is outside of the scope of this document. However, public records should not be exempted from the definition of Commercial Data.

[2] Privacy Impact Assessment Guidance, March 2006, P.10-11.

V. Sources of Commercial Data

The US commercial sector collects and processes personal information continuously. In virtually every action and transaction, some form of personal data is gathered, processed, resolved, and stored. Below are descriptions of various types of Commercial Data.

A. Financial Services

US consumers leave a trail every time they perform a transaction, for example, when they use credit and debit cards for purchases, conduct online banking, or use Web-based services for insurance claims.

The list of digital transactions in the financial sector is too long to enumerate here. It includes banking and insurance, the buying and selling of stocks, transferring currency, and buying, renting or leasing equipment, homes, and vehicles. Further, the commercial entity engaging in the transaction also has the ability to use the payment method to process PII pertaining to the individual.

Sources of this type of PII include banks, credit card issuers and processors, merchants, insurance companies and brokerage houses.

B. Communications Services

The increase in the use of communications technology has created a vast amount of telecommunications traffic. Each call is logged, tracked, billed and stored, creating a large data set. Sources of this type of PII include telephone companies, local and long-distance service providers, internet service providers, wireless services providers.

C. Location Data

Using electronic payment methods, telecommunications and navigation tools can yield even more information relating to the individual's location. Critical in an emergency, this information can be useful for a variety of other purposes (e.g. determining when individuals meet with each other).

Sources of this type of PII include financial, communications, and retail businesses

D. Interagency Exchanges

. Government agencies may obtain PII directly from non-government entities, and then may transfer it to another government agency. Due to the difficulty in determining the first source of the data, this paper applies to these situations.

VI. Methods of Access – The GAO has observed that the Departments of Justice, Homeland Security, State and the Social Security Administration had approximately $30 million in contracts to access Commercial Data in 2005.3 These agencies primarily used the Federal Supply Schedule of the General Services Administration contract and the Library of Congress's Federal Library and Information

[3] United States Government Accountability Office, *Report to Congressional Committees: Personal Information: Agency and Reseller Adherence to Key Privacy Principles*, GAO-06-421, p.4

Network to obtain access to the information. The uses of these data ranged from law enforcement activities to FEMA's detection of fraud in disaster assistance applications. Government agencies can access Commercial Data in a variety of ways:

1. Purchase

Government agencies can execute contracts to purchase or access Commercial Data

2. Direct Collection

Government agencies can ask government employees or contractors to use available tools (e.g. internet search engines) to access Commercial Data

3. Request

Government agencies can make voluntary or informal requests for Commercial Data from private entities.

4. Legal Process

Government agencies often have the ability to obtain Commercial Data through subpoena or through litigation discovery.

VI. Obligations Accompanying the Use or Acquisition of Commercial Data

Individuals have developed expectations for the responsible stewardship of PII that pertains to them. Indeed, several laws obligate businesses to implement information privacy and security procedures to protect PII entrusted to them. Many companies, taking their responsibilities for trusted customer relationships seriously, have promised to care for and protect PII they possess, even lacking specific regulatory obligations.

Commercial enterprises have PII about individuals sourced from other commercial enterprises. Credit card processors receive payment authorization requests from their subscriber companies, data processors receive PII for processing and analysis, and shipping companies regularly receive information with packages for delivery.

A chain of responsibility begins when a data collector gathers PII under a set of explicit, implicit, or regulatory obligations. Whether these obligations set relatively easy or difficult standards for the protection of the information, they are enduring obligations that should travel with the data and whose observance should be required regardless of who controls or processes the data. Government use of this PII brings the potential for increased adverse impact to the individual, and therefore should require even greater transparency and care.

VII. Uses of Commercial Data

One of the difficulties in providing guidance on the government's collection, use and/or storage of Commercial Data is the diversity of uses of the data. Some of these uses are:

A. Verifying Data

The government uses the Commercial Data to verify that data it has is correct. (e.g. a search of internet phone directories to verify an address, and the address is determined to be correct, therefore nothing is recorded).

B. Verifying Data and Storing Inferred Conclusions

This is similar to the use specified above, but here the government makes a record of the result of the verification (e.g. a government employee makes an entry in a database that an address is not verified, but does not enter any other information in the database). By linking these inferred conclusions to PII, the conclusions also become PII.

C. Bringing Data into a Government Database

The government enters the Commercial Data or inferred conclusions (beyond verification) into a government database, (e.g. a government employee could use a Commercial Data vendor to understand whether individuals applying for disaster assistance have poor credit history, and then stores the individuals' names in a database field entitled "fraud risk").

D. Using Analytical Tools

The government employee uses software, or has a vendor use software, to analyze data and produce a report. The government employee either stores the results of the report, causes a vendor to store the results, has the government take an action based on the results, or has a vendor or other entity take action based on the results. (e.g. a request to a Commercial Data vendor to use software to determine individuals who have lived in certain cities and have applied for certain permits).

E. Ad Hoc v. Systematic Uses

All of the above enumerated uses of data can be made on an ad hoc basis (e.g. a one time use of an internet search engine) or as part of a systematic program (e.g. a defined process that instructs government employees to use the internet search engine). There will be cases where it will be challenging for the DHS Privacy Office to determine whether activities are ad hoc or systematic (e.g. widespread and uniform use of a vendor, but where such use is not required by a program or process).

Bringing Data into Government vs. Having Vendors Store the Data: The government may collect, process and store Commercial Data (e.g. a request to a commercial enterprise to voluntarily provide data to the government) or the government may contract with a vendor to collect, process and store the data on the government's behalf.

VIII. Oversight of Agency Use of Commercial Data

A. Discussion of Agency Use

Because of the large amount of Commercial Data available, and the extent to which it can provide an understanding of the private life of individuals, the Committee has explored what oversight mechanisms are appropriate. The Committee's recommendations require the DHS Privacy Office to determine when individuals have a reasonable belief Commercial Data should be kept private.

Reasonable Belief that Data Will be Kept Private: The key analysis in determining what controls are necessary for DHS's access to or collection of Commercial Data should be whether the individual reasonably believes the data will be kept private. Two critical elements of this belief are:

a. Circumstances of Collection and Use – The circumstances of how the data were originally collected from the individual (e.g. what notice or commitment was made to the individual, or was the PII actually obtained without the individual's knowledge). As noted previously in this document, when an individual provides some PII (e.g. registering to vote, posting information to their own blog) they likely should expect the data may be used by others. However, when providing PII for financial transactions, an individual may have a reasonable belief that the data will only be used for a specified set of purposes.

b. Risk of Harm – How sensitive is the data? Could the individual be harmed by the use of the data (e.g. medical information, financial data, data concerning political affiliation or sex life).

Individuals may have a reasonable belief that data will be kept private regardless of whether the use of data by DHS is ad hoc or systematic; or whether the data is brought into the government or whether the data resides at a vendor. DHS will likely have many ad hoc collections of Commercial Data, that should not give rise to additional controls (e.g. an agency employee updates her contacts folder). Therefore, the Committee concludes additional controls are necessary where the collection, use or storage of Commercial Data is systematic or where the ad hoc collection, use or storage poses a substantial risk of harm to the individual.

The Committee has analyzed the following oversight mechanisms:

B. DHS Privacy Office Review of Contracts

The DHS Privacy Office can serve a useful role in reviewing the relationships DHS establishes with Commercial Data vendors, and ensuring the contracts for such relationships contain adequate protections to safeguard individuals' privacy. The U.S. government currently mandates privacy language in relevant procurement contracts. However, the Committee believes the current template language is unclear in whether it will apply to some situations of the use of Commercial Data, and may not include robust enough provisions. Therefore, the Committee recommends the DHS Privacy Office analyze the current language, and propose, if necessary, additional template privacy language for the contracts with such vendors and drive the implementation of that language through the relevant procurement organizations. The DHS Privacy Office should also analyze how it can ensure the vendors are fulfilling these contractual obligations.

C. System of Records Notices

The Privacy Act provides requirements for the completion of systems of record notice (SORN) when data exists in a system under the control of a government agency from which information is retrieved by any identifying particular or when a contractor does the same on behalf of the government.

Many instances of the use of Commercial Data will arguably fall outside of the requirement for the filing of a SORN, as the Commercial Data vendor operates their business for a variety of customers, not just the government. Also, for some government agency uses of Commercial Data, a Privacy Act exemption may apply, such as the exemption for records compiled for law enforcement purposes. As the SORN provides the best transparency to the public of the data processing, the Committee recommends the filing of a SORN for every DHS use of Commercial Data which is either systematic or has substantial risk of harm to individuals.

D. PIA Module

The Privacy Impact Assessment (PIA) is required by the E-Government Act of 2002 for systems developed after 2002 or a system undergoing a substantial change. OMB's guidance for the completion of the PIA requires a PIA when agencies "systematically incorporate into existing information systems databases of information in identifiable form purchased or obtained from commercial or public sources."4 The OMB document further states that "merely querying a database on an ad hoc basis does not trigger the PIA requirement." DHS has the authority under the Homeland Security Act to expand the situations where an agency component must complete a PIA.5 The DHS Privacy Office requires a PIA when commercial data is "used in a decision making process,"6 but not necessarily when the queries are only made on an ad hoc basis.

The DHS Privacy Office's current guidance document on PIAs refers to the collection of Commercial Data. Section 1.2 states (emphasis added):7

1.2 From who is the information collected?

1.2.1 List the individual, entity or entities providing the specific information identified above. For example, is the information collected directly from the individual as part of an application for a benefit, or is it collected from other sources, such as commercial data aggregators?

---

[4] OMB, *Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002,* Memorandum M-03-22 (Washington, D.C.: Sept. 26, 2003).

[5] Section 222(1) of the Homeland Security Act requires the DHS Chief Privacy Officer to assure "that the use of technologies sustain, and do not erode, privacy protections relating to the use, collection and disclosure of personal information."

[6] Written statement of Maureen Cooney Acting Chief Privacy Officer and Chief Freedom of Information Act Officer U.S. Department of Homeland Security before the joint hearing of the Subcommittee on Commercial and Administrative Law and the Subcommittee on the Constitution, Committee on the Judiciary of the U.S. House of Representatives, April 4, 2006.

[7] Privacy Impact Assessment Guidance, March 2006
http://www.dhs.gov/interweb/assetlibrary/privacy_pia_guidance_march_v5.pdf

1.2.2 Describe why information from sources other than the individual are required. For example, if a program is using data from a commercial aggregator of information, state the fact that this is where the information is coming from and then in 1.3 indicate why the program is using the source of the data.

Appendix I exempts ad hoc collection in the following section:

Commercial Sources

when agencies systematically incorporate into existing information systems databases of information in identifiable form purchased or obtained from commercial or public sources (Merely querying such a source on an ad hoc basis using existing technology does not trigger the PIA requirement).

The Committee believes that defining systematic access is difficult in practice. Whether a particular use of data has the regularity to be systematic is more difficult than determining whether the access is just a query as opposed to incorporating data into existing databases. Because of the difficulties in determining what is a regular and systematic access to Commercial Data, and what is "ad hoc" access, and because ad hoc access can still result in harm to an individual, this document recommends that a DHS program conduct a PIA (which should include a specific module on Commercial Data) for any systematic collection, use or storage of Commercial Data, and where ad hoc collection, use or storage has a substantial risk of harm.

Below is a list of topics the DHS office should analyze thoroughly in the PIA. Most of these topics are currently addressed by the PIA, but whether the PIA serves as an effective oversight mechanism will depend on whether DHS provides adequate staffing to manage the PIA review process, and whether there is a management mechanism in place to ensure individual organizations cooperate with the Privacy Office to allow for a robust review of these topics. Further, privacy protection requires more than the processing of forms. The Privacy Office must be given sufficient authority to require DHS organizations to modify their programs and supporting technology solutions to mitigate privacy risks. The Privacy Office must be able to act on this authority without feeling such action could negatively impact their ability for effective action in the future. The Committee recommends expanding Section 1.2 of the PIA to specifically call out the following for each collection or processing of Commercial Data. Some of these categories overlap with other sections of the PIA. It is important, though, for this "Commercial Data" module to call out this information for a separate thorough analysis, due to the increased risks from the use of Commercial Data.

1. Purpose

What is the legal basis for accessing or obtaining the information?, What is the specific reason why the government agency needs to access or obtain the Commercial Data?, What is the specific use of Commercial Data being proposed?

2. Onward Transfer Obligations

Do pre-existing obligations for the handling of the data flow to the government?

3. Scope of Request

Is the requested data the minimum amount necessary to serve the purpose for which the data are being accessed or collected? How has the agency minimized the use of Commercial Data to access or collect only that data which is required for the specific Purpose?

4. Method of Access

What is the method the government uses to access or collect the data: request, purchase, direct collection using the internet or other tools, subpoena, obtained through another government agency?

5. Efficacy

Is there a less intrusive means to achieve the program objective than using Commercial Data? (e.g. can the same purpose be achieved by using data already held by the government).

6. Transparency

If possible, the government agency accessing or collecting the data from the commercial source, should ensure that the commercial source has provided notice to the individual that the data may be accessed by or transferred to and used by other entities (e.g. DHS). Also, the government agency should when possible provide effective notice of the processing of the data. The analysis of the PIA should consider the following questions: Will providing notice to individuals about the access to or collection and use of Commercial Data frustrate the specific purpose? Is the notice provided in a way that individuals are likely to understand how the data are processed?

7. Data Quality and Integrity

Commercial Data have a varying degree of data quality and integrity. This may have limited impact on an individual when used for marketing purposes. If government agencies will use these data for security or law enforcement purposes; however poor data quality could potentially lead to substantial harm to an individual (e.g. imprisonment, suspension of rights to travel). The PIA should scrutinize the extent to which the data quality of the Commercial Data has been reviewed and whether it is sufficient for the agency's use.

- Has the Commercial Data's quality been measured as appropriate for the specific purpose?
- Has there been a balancing of potential benefit and potential harm from accessing the data?(The level of sensitivity to the individual of the data collected should be a factor in this analysis)
- What process will be put in place to manage individual complaints that information in the data set is not accurate?

8. Unintended Consequences

Does the use of the Commercial Data create a weak link in the security chain? The use of Commercial Data for security and law enforcement purposes could actually decrease security. If Commercial Data

sources' internal security procedures (e.g. background checks of employees) are less secure than the government's, then these sources will become the weak link in the security chain. This weak link could create an opportunity for organized crime or a terrorist network to have an operative infiltrate the company and either destroy damaging data or modify data to implicate other individuals.

9. Limited Use

How is the retention and sharing of the data being controlled to meet indviduals' reasonable privacy expectations?

10. Control Mechanisms

The agency review of the PIA should make certain the following items are addressed sufficiently.

What provisions to protect the Commercial Data will be placed in the contract with the data vendor?

- Training
- Access controls
- Use policies
- Access and use logs
- Audits

11. Oversight

The DHS Privacy Office must be given the necessary resources to adequately review the PIA, and the ability to freely audit compliance with the statements made in the PIA. Further, they must have sufficient authority and freedom to ensure the requirements of the PIA are met and any issues are remediated. If required processes are not followed or issues are not remediated timely, there should be substantial consequences for those officials directing the processing of the data. Further, the PIA should address how the government agency is providing oversight to ensure the following issues are managed appropriately.

- Separation of duties (e.g. managing access rights to the data to decrease inappropriate use of the data, such as making certain an individual cannot both act as their own approver for the provision of benefits)
- Making certain control mechanisms have sufficient authority to protect against misuse of the data (e.g. privacy officers in an agency component must have authority to require changes to a program).
- Incident reporting, response, and learning
    - Unauthorized access and/or use
        - Internal
        - External

- Data corruption, loss or compromise
- Incident capture, tracking and escalation
- Investigation and remedies

12. Lifecycle

Is there a formal process in place for requests for access, deletion and redress?.

IX. Conclusion

Government has a unique ability to restrict individual rights, such as the power to imprison. Due to this increased risk of harm to individuals from government's collection of, access to and processing of Commercial Data, the Committee believes the recommendations contained in this document are warranted. The Committee has focused its recommendations on the existing PTA and PIA processes to minimize the administrative burden on DHS and the DHS Privacy Office. The Committee seeks input on these recommendations, and may respond to comments by either revising this paper or publishing additional analysis.