

Report No. 2011-_____

Privacy Policy Recommendations for DHS Federated Information-Sharing Systems

This white paper reflects the consensus 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 or Department). The Committee's charter under the Federal Advisory Committee Act is to provide advice on programmatic, policy, operational, administrative and technological issues within DHS that relate to personally identifiable information (PII), as well as data integrity and other privacy-related issues.

The Committee deliberated on and adopted these recommendations during a public meeting on _____, 2011, in Washington, DC.

Privacy Technology Guidance

A. Controlling Access to a Shared Database

Program decisions on the degree to which access controls should be centralized will be critical points for the DHS Privacy Office to provide input and guidance to the creation of a federated information sharing system. Federated access control systems contain many of the same issues as federated identity management structures and considerable guidance can be derived from the work done in the past in that area, and work currently being done on the National Strategy for Trusted Identities in Cyberspace program www.nist.gov/nstic.

Each federated database that contributes data to a DHS federated information sharing system will likely determine the classification of its data and prescribe rules on entities or individuals who should not access and/or receive the data. A centralized access control system will be necessary, especially given the lack of a full understanding of the potential uses of the federated data or of the classes of entities who may gain access such as non-DHS federal agencies and state/local/tribal organizations. From a process standpoint, the access control rules will have to be specifically delineated and made fully operational in corresponding technology solutions. Moreover, different account types will need to be identified, the conditions for group membership established, and access to the federated information sharing system should be predicated on specific conditions. These conditions should include multiple, auditable access control mechanisms, incorporating a variety of attributes important to the organization (e.g., role, intended use, physical location, case assignment), appropriate to the requested data, and to the source systems.

The Privacy Office will need to have dedicated resources to help both guide the creation of this mix of centralization and federation, and to provide oversight of the regular risk assessment as to whether the system is behaving appropriately. Additional access control systems and processes will be required to be put in place for the log data created to provide reasonable security and accountability. A determination must be made at the outset over whether the Privacy Office should operate these systems, establish a technical reporting system for their operation, provide general assessment and oversight, or some combination of these roles. An explicit determination of who will take on these roles, and why, should be determined early in the design phase. In any case, the Privacy Office should play a central role in the development, testing, deployment and oversight of the system's design, function, and operation.

B. Data Integrity and Quality Assurance

The issue of interoperability of the data structures of the federated databases requires immediate technical attention, with guidance from the DHS Privacy Office. It is unlikely the data in the source databases are currently stored in a manner that allows for easy and accurate data relationships among them. DHS may need to create a template middleware translation to allow for similar, but different, data fields, formats and values to be combined.

In accordance with well-known business warehouse architecture concepts, attention will need to be paid to the major data layers of the system - data acquisition, data storage, and data presentation. Each of these layers plays an important part in assuring integrity and quality. For example, the data acquisition layer may address data in the different source databases and either load them into a data warehouse or 'normalize' the data to prepare it for queries. A main challenge may likely be the use of different attribute values across the different source databases. Data cleansing rules will need to be created to recognize the relationships between different types of data and to ensure their accuracy. The storage system for these data cleansing rules will itself need appropriate access control management processes and audit structures.

Another significant challenge will be to the need to automatically identify and resolve data conflicts that flag quality issues (e.g., two systems reporting different dates of birth for the same social security number). These conflicts will need to be logged and then communicated to the systems of record for resolution, which will also have to be overseen. Much of this process may be manual and may have privacy implications for individuals (e.g., determining which birth date is correct). This process may also present significant cost implications for the government, so prior similar efforts, both in DHS and elsewhere, should be analyzed before undertaking this effort.

An additional requirement will be the development of machine and system readable metadata tags and rules that would enable the management of data utility against policy requirements. For example, stale data may not be reliable for certain applications or functions. Likewise, confidence in the quality of certain data may be an attribute relevant to certain uses. As systems are designed (and redesigned) for future integration with the federated information sharing system, attention needs to be paid to meta-data tagging associated with data types,

data elements, data sources, data time-stamps, data retention periods and other factors that may be material to the reliability and quality of data for particular purposes. This is distinct from data accuracy, in that accuracy *per se* does not necessarily address relevance and fitness for specified uses. Policies need to be developed addressing the appropriateness of data. This is an area where the DHS Privacy Office can contribute by developing policies and review and approval processes to manage this aspect of data.

C. Redress

During the requirements definition phase for any resulting system, it is important to address the opportunity for automated redress within the context of the system of records. The DHS Privacy Office will need to work closely with the program management team to collaboratively develop requirements that ensure inclusion of redress goals. At a minimum, there should be one level of redress required for implications to the individual from the results of the centralized query to the federated information sharing system. However, to the degree the redress request requires an update to the system of record, there should be an automated way to process that request such that it reveals from which systems of record the original data came. Effectuating this redress mechanism will likely require the centralized database to understand and log from which systems the initial data came. The inclusion of this data in the centralized system will create additional access control and security requirements for that centralized log.

D. Secondary Uses and Onward Transfers

Because the system will provide responses to specific queries, users will draw inferences from the combined data. This is a stated goal of all federated systems and drives the need to have a mechanism to make certain the queries to the federated databases, and the use of the resulting inferences, do not violate privacy commitments made by the source system of records. It is unlikely such a mechanism can be manual, so a serious design effort for an automated system must be undertaken. These commitments will include representations made in Privacy Impact Assessments and Systems of Record Notices for Federal Systems, but may also include policy commitments, state/local/tribal laws and published privacy policies. Setting up the centralized mechanism for transferring these requirements from the systems of records, matching them with the proposed new uses or transfers, and maintaining auditable logs to understand how the decisions will be made, will be a significant undertaking. This work will require substantial resourcing from the DHS Privacy Office. There are other efforts within the Federal Government actively addressing these kinds of requirements in software systems and those technology developments should be leveraged herein. The DHS Privacy Office will also need to work with their component privacy professionals who have oversight responsibility for the underlying systems.

E. Applicable Privacy Policies and Standards Development

The federated information sharing system should have a machine readable privacy policy to help manage secondary use and onward transfer and other privacy management requirements. There is considerable history in machine readable privacy policies and their technical implementation. DHS should consider mandating the use of machine readable privacy policies for the databases that will comprise the federated information sharing system.

Work is currently underway in the standards development community, including OASIS, ISO/IEC and other recognized standards bodies, to develop standards that can be used to build automated implementations of privacy management controls. The Privacy Office technology staff should explore engagement with this important work in standards to inform the process and help provide use cases applicable to DHS needs. Additionally, the Privacy Office should coordinate with other government agencies, such as NIST, while also recognizing the additional value that may come from direct engagement in the standards process. Ultimately, DHS should determine the appropriateness of adopting specific standards applicable to its systems. Contributing to the standards development process can also help drive technology innovation and the integration of the standards into commercial, off-the-shelf products

F. Accountability

For the appropriate oversight personnel (over both the information sharing system and the federated source systems) to be accountable for the commitments described above, it will be necessary to provide the technical ability for them to perform periodic risk assessments of the system and to understand the results of those risk assessments (if they have oversight responsibility of one of the federated databases). Tools will need to be developed, or acquired, to provide oversight officials with the appropriate access and log data to assess the system.

Technologies that support governance, risk management, and compliance (often referenced collectively as GRC) should be integral components of the federated information sharing system. While distinct, these three GRC components are inter-related and their integration within the federated architecture, system design and operational reporting systems will enhance oversight and visibility into the overall system and its trust posture.

Governance tools are critical because they will support the development and management of organizational policy requirements and the chain of control needed to ensure oversight, management awareness and remedial action. In the federated information sharing environment, high-level management and oversight are critical to ensure privacy and public trust in the system. Similarly, risk management and compliance controls and supporting technologies are critical for meaningful, ongoing oversight of the system once operational. Technical and personnel components are in constant flux, and reside in an ever-changing threat

landscape. Risk management and compliance technical tools will enable appropriate insight into risks, risk management adjustments, and compliance reporting.

It will be important for the DHS Privacy Office to have formal points to engage in the further development of this system. Further, there will be value in this Committee re-engaging at the point when formal requirements with traceability to specific governing policies and regulations are specified, and when major system development or acquisition decisions will be made. The Committee can also provide additional value at the points where major policy decisions will be made, and when the audit tools are being developed.

G. Audits of System Usage

Logs should be kept in a data warehouse that can be queried and used to generate reports. This will also allow searching for patterns through data mining to shed light on the who, what, when of a data access event, and also potentially how, when and with what other data it is being integrated. Automated tools are available and should be used to carry out these pattern searches and reporting on anomalies should be done on a close-to-near-real-time basis. Using automated tools will allow for fewer people to view the personal data, and may thereby be privacy enhancing. In addition, these automated tools should be designed to automatically detect privacy issues (deviations from obligations) by testing queries access of data in the source systems against defined rule sets.

These tools need to account for issues that may arise from classified queries. If the individuals who are performing audits of the source systems do not have authorization to access the classified query, then DHS needs to provide appropriate protection of this classified information while also preserving the integrity of the source system audit.

H. Data Retention

The data retention policies for the information sharing system should be predicated on the following two principles. 1) The actual queries (not the data retrieved therefrom) should be saved for the longest regulatory period, so that audit logs can be effective in understanding what people are querying and why; 2) The data inferred from those queries should be saved for the shortest regulatory period possible (essentially consistent with the reason why that query/data was assembled in the first place).

I. Data Security

Assuming that a DHS information sharing system will support long-term storage of aggregated data, this system creates an attractive centralized target for malicious actors. The appropriate level of baseline controls are those specified in NIST SP 800-53 for high-impact systems where the baseline is set to protect against threats from highly skilled, motivated, and well-resourced threat agents. Due to the aggregation of clearly sensitive data from multiple sources, the security controls implemented within this system must be high, and the program managers will need to work closely with the DHS information security staff, while continuing to seek input from the DHS Privacy Office to ensure that the selected security controls are appropriate. The

Privacy Office should work with the appropriate DHS security offices to develop continuous monitoring policies. They should also review business process requirements and establish reporting instruments with respect to an effective continuous monitoring regime.

DRAFT