



Homeland
Security

Privacy Office

Protecting privacy while promoting transparency



Data Privacy & Integrity Advisory Committee

Public Meeting

Thursday, January 30, 2014

2:00 - 4:00 PM



Homeland
Security

Privacy Office

Web Conference Instructions

Please follow these instructions:

CONFERENCE LINE

- Dial **800-619-0355** and enter passcode **4386646**.
- Please mute your phone but don't place it on hold.

QUESTIONS

- Hold questions until the end of each session when the operator will open the line. DPIAC members have priority.

HANDOUTS

- This presentation is also available on our website: www.dhs.gov/privacy. Click on *Events*, then *DPIAC Meeting Information*.



**Homeland
Security**

| Privacy Office

DHS Privacy Office Update

Karen Neuman, Chief Privacy Officer

DHS Privacy Office Accomplishments:

- Staffing
- Outreach
- International
- Compliance
- FOIA
- Oversight



**Homeland
Security**

| Privacy Office



DHS Data Framework

Rebecca Richards
Senior Director, Privacy Compliance

Agenda

- Review mission requirement
- Update on status
- Answer questions
- Review DPIAC Tasking

Context

- DHS data is gathered under a wide variety of legal authorities
- DHS legal, policy, privacy, and CRCL requirements complex
- Multiple domains with a variety of user populations
- Multiple data formats and standards



DHS NEEDS
TO KNOW
WHO THEY
KNOW...
RIGHT AWAY

The Mission Challenge is:

- Operators and leadership manually check numerous data holdings across many DHS components
- Operators spend significant effort determining if resulting records are actually the same person
- Current approaches do not address Privacy and legal restrictions, limiting ability to aggregate data



DHS Data Framework Long-Term Operational Impact

Advanced access controls give data stewards the confidence to provide data for community discovery

Significantly faster ability to discover, access, search, and exploit relevant information



Fine-grained access to data and enhanced audit logging greatly reduces possibility of security breach

Enhanced analysis capability achieved through a more complete picture of the Department's and IC data

Enhanced search and discovery capability on complete mission data ensures comprehensive and repeatable search results

High Level Approach

- **Guiding Principles:**
 - *Enable scalable and controlled* aggregation of DHS datasets
 - *Design built-in safeguards* for access and use of DHS data
 - Proven *governance* process
 - *Drive new analytics* to enhance efficiencies and mission capabilities
 - *Enhance Privacy Protections* and Safeguard the data.
- This initiative will be informed by the current pilots underway demonstrating the ability to control and safeguard DHS information, while supporting our operational need for advanced analytics.

Key Elements of DHS Data Framework

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 (i.e., authorized function), with the purpose for which data can be used (i.e., authorized purpose); and
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.

DHS will log user activities to aid audit and oversight functions

Major Pilot Activities

- 1. Neptune Pilot** Data will be tagged and ingested in a “big data” platform on the SBU domain. Data in the Neptune Pilot will be shared with the CEI Prototype and the Cerberus Pilot, but will not be accessible for other purposes.
- 2. CEI Prototype** The CEI Prototype, residing on the SBU domain, will receive a subset of the tagged data from the Neptune Pilot and correlate data across component data sets.
- 3. Cerberus Pilot** The Cerberus Pilot, residing in the TS/SCI domain, will receive all of the tagged data from the Neptune Pilot and test the ability to ensure that only users with certain attributes are able to access data based on defined purposes using the rules engine.



Three Data Sets in the Pilot

- 1. TSA Alien Flight School Program (AFSP)*
- 2. CBP Electronic System for Travel Authorization (ESTA)*
- 3. ICE Student Exchange and Visitor Information System (SEVIS)*

DHS Data Framework

DHS SOURCE SYSTEMS

DATA TAGGING & STANDARDIZATION

PILOT SYSTEMS



Tagging Engine

Landing Zone

NEPTUNE
(Implemented)



Ingest & Correlate



CERBERUS



CEI
(Implemented)

- Selected DHS Data Sources
- All Data elements
- Un-correlated data

- All DHS Data Sources
- 9-11 Data elements
- Correlated / Indexed data

Privacy Compliance Documentation

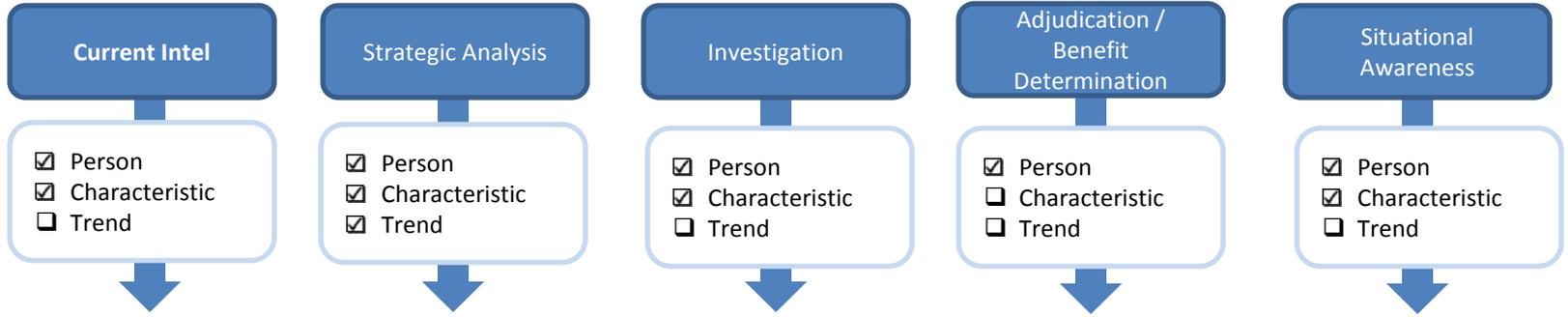
- System of Records Notice
 - Common Entity Index Prototype published August 23, 2013
- Privacy Impact Assessments published November 6, 2013
 - DHS/AII/PIA-046 DHS Data Framework
 - CEI Prototype
 - Neptune
 - Cerberus

Pilot Activities

- Tag the data (Neptune)
 - Core, Extended, Encounter Data
 - Source system information.
- Control Access and Use (Cerberus and CEI)
 - **Function** - what you are doing
 - **Purpose** (+ *attributes*) determines which **data sets** and **categories** you can search against
- Match entities across data sets (CEI)
- Create Immutable Audit Logs (Neptune, CEI, Cerberus)

Controlling Access and Use

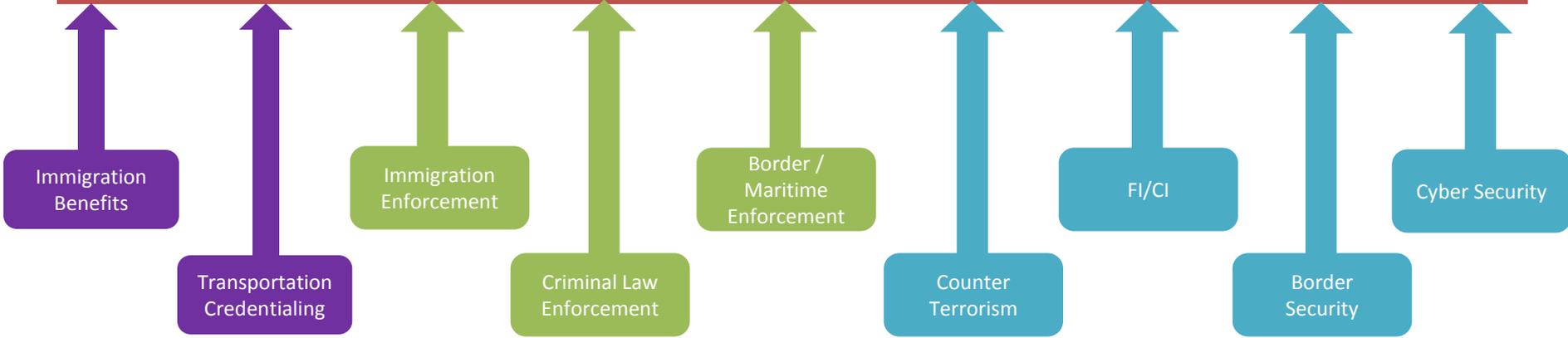
The **function** you're performing (that is, **what** you are doing)...



.....determines what **type of search** you can perform

Search type **ON** data sets

.....(+ attributes) determines which **data sets** and **categories** you can search against



The **purpose** that you're working toward (that is, **why** you are doing)...

Access Control Demonstration

- Access control through the use of data tags, context (function + purpose), and user attributes
 - Visibility within a dataset:
 - Users can only see their component data, ensuring protected externalized authorization
 - Ex: AFSP user can only see AFSP data and not ESTA/SEVIS data
 - Visibility across datasets:
 - “Super user” with full access
 - Users with partial or limited access
 - Ex: User who can see ESTA/SEVIS, but not AFSP
 - Ex: User who can see core and extended data, but not encounter data
 - Administrative access:
 - System administrators have ability to update or change users’ authorization but do not have access to data or analytical tools
- Search Capability
 - Name or Character Searches:
 - Joseph A. Smith, DOB 3/29/70
 - John or Joseph Smith, male, Canadian Citizen, arriving at JFK on DD/MM/YYYY
 - User can view the information in a clear and accessible format
 - Baseball Card Widget

Mitigating Privacy Risks

- In order to move from pilot to operational, the following must occur:
 - A clearly defined **leadership and governance structure** for the ongoing development and maintenance of these systems which incorporates both operational and oversight components
 - Must include clear criteria for approving additional data sets, additional uses, and new analytic tools
 - Improved **transparency to the public** at the point of collection, in the applicable SORNs, and PIAs for data being stored these systems
 - Ability to push **timely updates** from operational systems to Neptune/Cerberus/CEI.
 - Defined process for providing access and **redress**
 - Defined process for **auditing the immutable logs** and development plan for using technology to identify unusual or anomalous behavior in the system
 - **Training for users** on the appropriate use of the system, the data and where to get more information on it, and how the system will identify possible misuse of the system.
- Pilots demonstrated the following:
 - Demonstration that **data tagging** is occurring properly in Neptune
 - Demonstration that the **dynamic access controls** work in CEI and Cerberus

Neptune

Tagging the data

- Tagged data with source, core, extended, and encounter.
- Transferred data to CEI Prototype and Cerberus
- CEI Prototype did not receive any encounter information.
- CEI Prototype did receive core and key elements of extended biographic
- Cerberus received and used the tags successfully.

Next Steps

- Developing approach to provide timely update from operational systems to Neptune
- Identified areas for improvement with the quality of the data at the source

CEI PROTOTYPE SUCCESS

✓ *Use real data*

- Data from source systems (**CBP-ESTA, TSA-AFSP and ICE-SEVIS**) was **tagged and transferred** to CEI for **correlation and discovery**

✓ *Ensure protected externalized authorization (access control)*

- DHS users **logged-in** to the CEI environment for **authentication** and create a query on an individual of interest
- CEI automatically **checked the users' attributes** and determined which data sources and data types were available to the user
- CEI **returned all results that were relevant to the query and permissible** based on the users' attributes

✓ *Provide correlation value*

- **Data from multiple data sets was correlated** within CEI based on correlation rules
- CEI search results **returned pre-correlated data (identities)**

TRUST



VALUE



Cerberus

Secure data cloud environment on the TS/SCI network that provides:

- Consolidated data repository of **unclassified and classified** data from DHS and the IC
- Controlled data service to enable **classified queries, discovery, dissemination, and analytics**
- Policy-based dynamic access control implementing DHS Data Framework user attributes, defined data tags and contexts
- Data **protection** and use **auditing**
- Data quality and retention management

Access control must be proven before operationalizing!

Anticipated Cloud Capabilities

ANALYTIC TOOLS

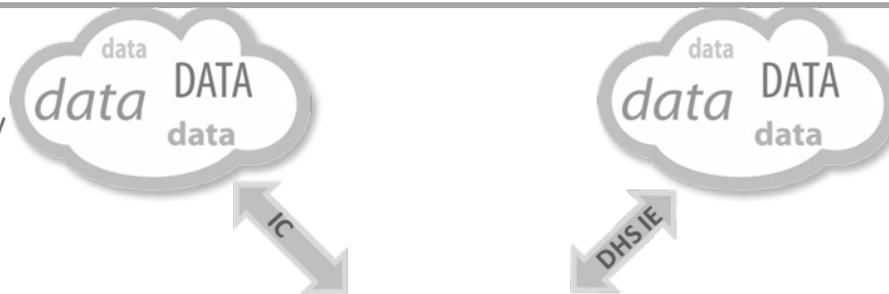


Innovation Zone

Data Insights, Mapping, Visualization

Analytics and Discovery

CERBERUS AS A CONDUIT/
STAGING AREA

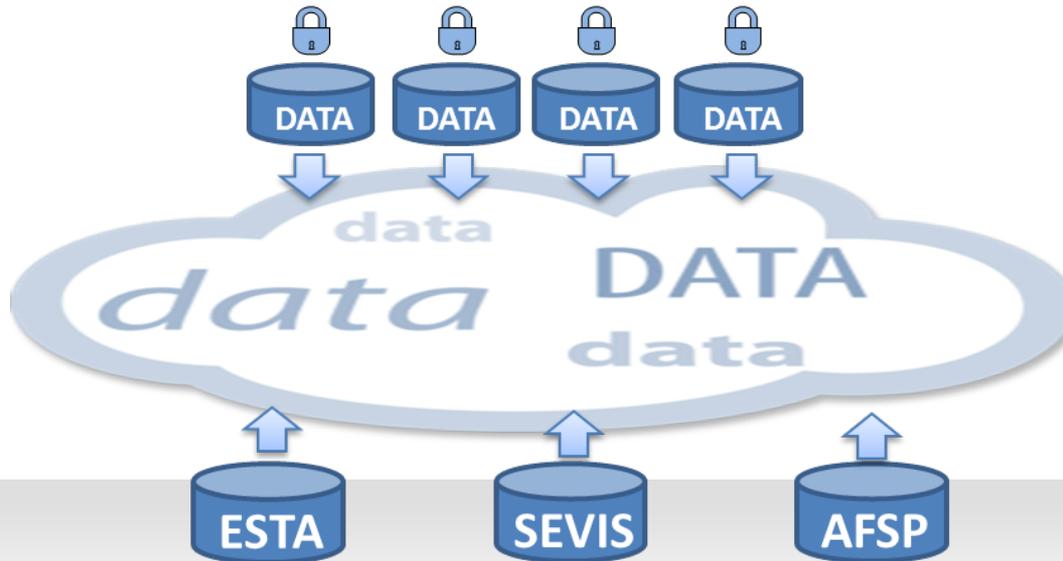


Enable DHS IE

Share data with the IC
via Cerberus

Participate in IC Clouds/IC ITE

CERBERUS AS A
"DATA LAKE"



Access Control Policy

Additional Data
(Components, IC, DOUGLAS)

"OneSearch" Capability

Initial Data Sources

What are we learning from the pilots?

- Brought people from different parts of DHS to work together in new ways.
- Allowed us to test and prove concepts that had never been implemented at the enterprise level
- Comingled data (different sources, different authorities, etc.) in a single location
- Access to data is protected based on controls approved by both oversight and component data owners
- More data is accessible while being better protected
- There is a visible path forward for performing correlation across all relevant DHS data sets and faster time to analysis for mission operators

The processes, policies, technology, and organizational capability that allows the Government to detect relationships between people that DHS knows across components

Next Steps

- Finalize governance
 - Refining access control
 - Establishing policy and governance processes
 - Continue Stakeholder Engagement
- Continue planning for the next iteration of CEI and Cerberus
 - Demonstrate effectiveness of controls
 - Continue to mature Data Framework
 - Develop Mission CONOPS
 - Continue Stakeholder Engagement

DPIAC Tasking Request

- Notice/Transparency
- Auditing/Oversight

Public Comments: 3:45 – 4:00



**Homeland
Security**

| Privacy Office



Homeland
Security