



Big Data for Government Symposium

<http://www.ttcus.com>



@TECHTrain



TTC™

Technology Training Corporation

Linkedin/Groups:
Technology Training
Corporation



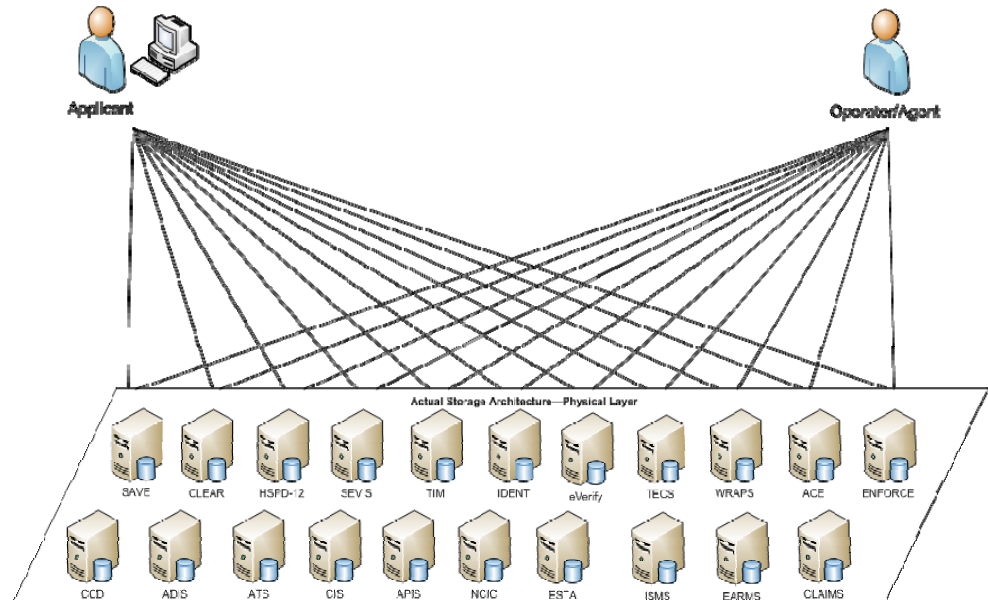
Big Data Strategies

Government Big Data Symposium

June 17, 2014

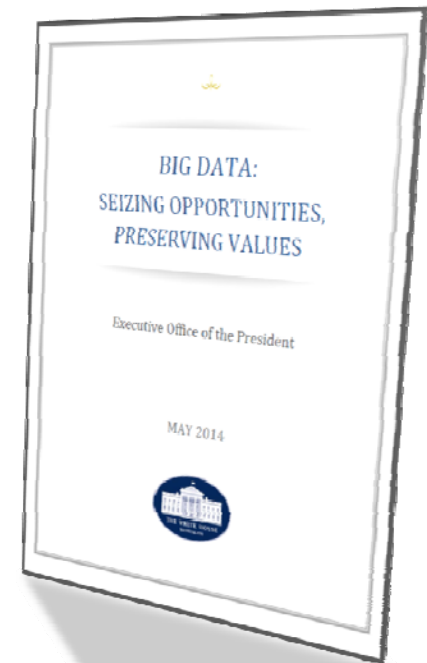
Mission Problem

- Today's mission constraints
 - Stove-piped systems and *inaccessible data*
 - *Multiple* log-ins
 - Different systems *deliver different search results*
 - Difficult and *inefficient use and sharing* of available homeland security information



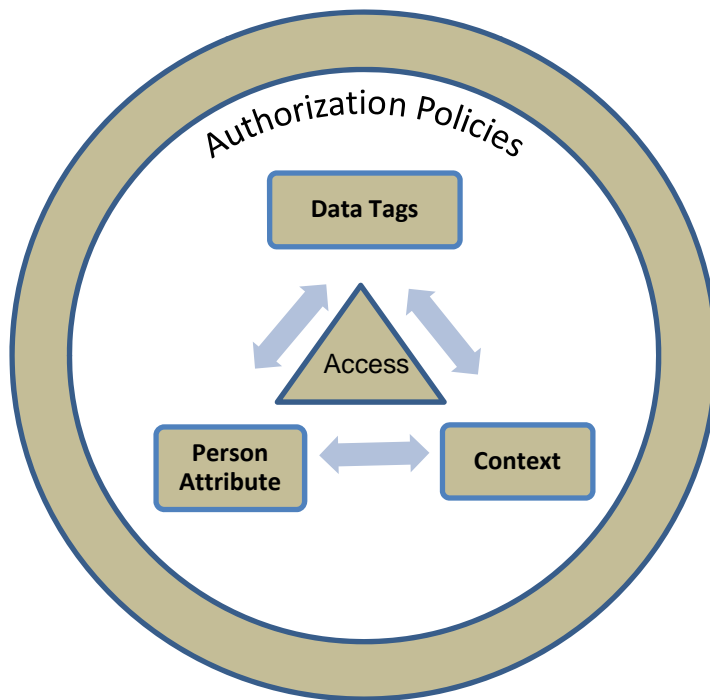
High Level Approach

- 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
- This initiative continues to be informed by the current pilots underway demonstrating the ability to control and safeguard DHS information, while supporting our operational need for advanced analytics.
- Called out in May 2014 Big Data Report as best practice



Progress Made to Date Controlling Access and Use

Access Control is the process of controlling the flow of data by making decision requests and authorization decisions based on policies.



Authorization Policies are the rules by which persons are granted or denied access to data or resources based on the alignment of their *personal attributes*, the *context* of their request, and the *type of data* they are requesting.

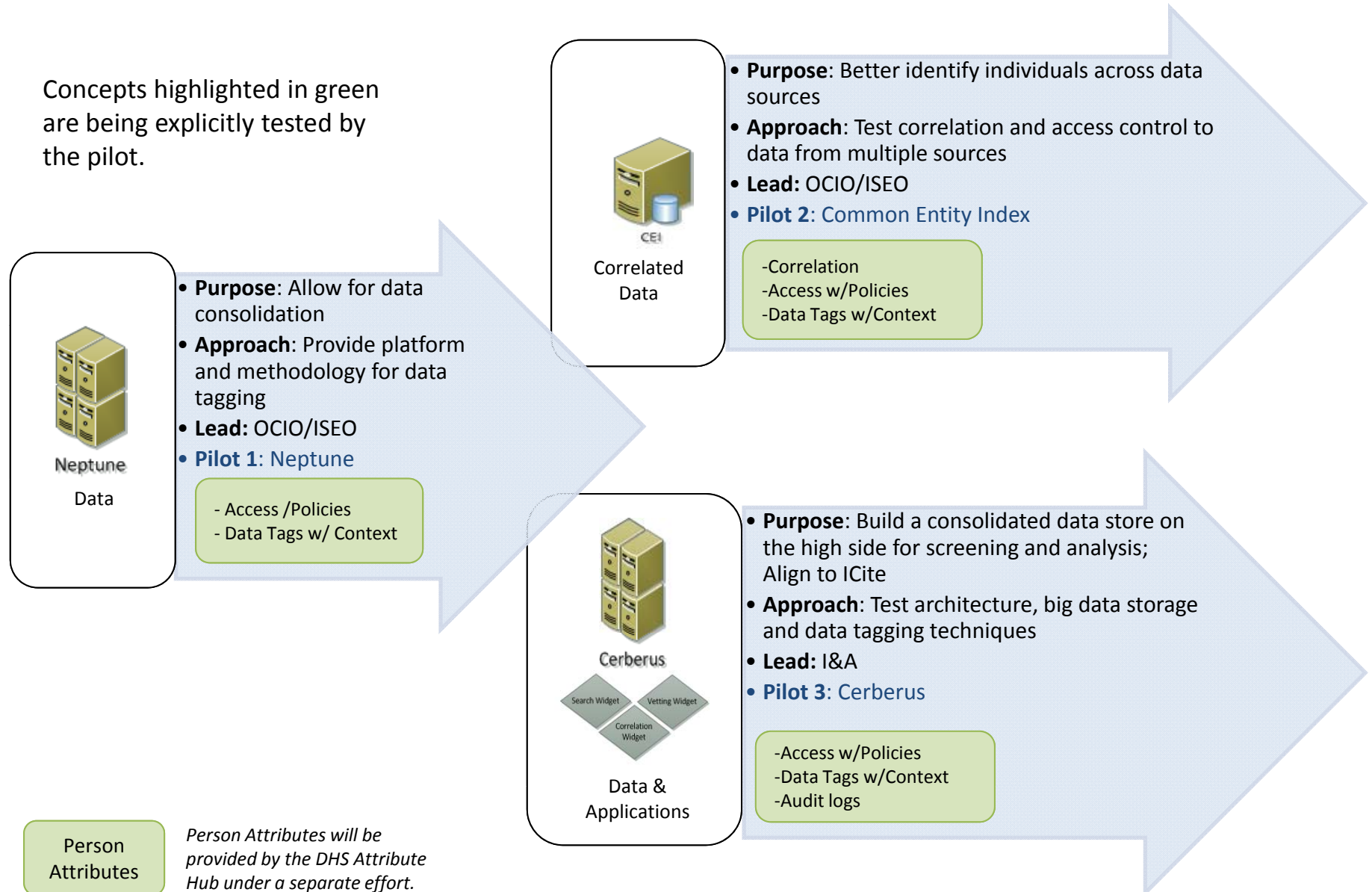
Person Attributes (User, Requestor) are characteristics of an entity requesting an operation on an object.

Context or Use is the purpose for which the data will be used (e.g. Law Enforcement) and/or the type of search/query conducted (e.g. Person search by name).

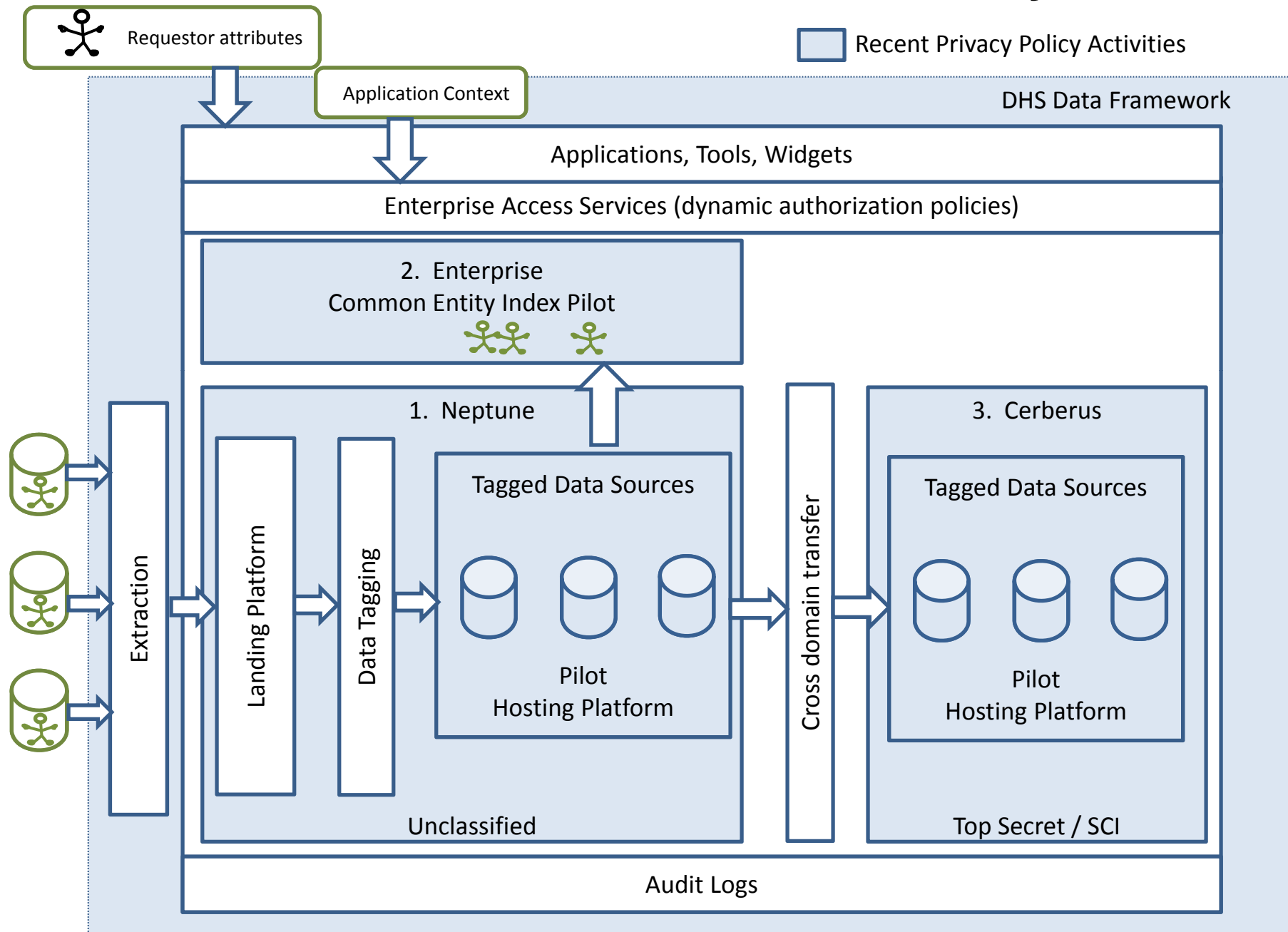
A data tag is characterizing metadata (data about data) associated with a data object. A data tag is both the tag name and the value, e.g., system name "ESTA".

Pilots and Relevance to Access Concepts

Concepts highlighted in green are being explicitly tested by the pilot.



2013 Pilot Current Architecture & Privacy Activities



Privacy Compliance Documentation

- System of Records Notice
 - Common Entity Index Prototype published August 23, 2013
<http://www.dhs.gov/sites/default/files/publications/privacy-pia-dhs-wide-cei-pilot-09262013.pdf>
- Privacy Impact Assessments published November 6, 2013
 - DHS/AII/PIA-046 DHS Data Framework
<http://www.dhs.gov/publication/dhsallpia-046-dhs-data-framework>
 - CEI Prototype
<http://www.dhs.gov/publication/dhsallpia-046-2-common-entity-index-prototype>
 - Neptune
<http://www.dhs.gov/publication/dhsallpia-046-1-neptune-pilot>
 - Cerberus
<http://www.dhs.gov/publication/dhsallpia-046-3-cerberus-pilot>

Key Challenges

- Data Stewardship / Governance:
 - Clearly defined leadership and governance structure required.
 - Dynamic access controls key to trust.
 - Demonstrations for data tagging, access controls and widget deployment were critical.
 - Oversight involvement critical: Privacy, CRCL and Legal
 - Review existing SORNs to increase transparency.
 - Ensure clearly defined process for access and redress in Cerberus and Neptune.
 - Defined process for reviewing immutable logs.