WHAT DOES BIG DATA MEAN TO DHS

Lots of Hype with lots of approaches

Lots of structured data that can be aggregated

Potential for uncoordinated mission initiatives and proliferation of more siloes

Finding the *right approach* to meet DHS Needs
WHAT BIG DATA CAPABILITIES TO USE

Focus is on meeting Mission Information Needs

Gathering the right data from the right sources

Data analytics specific to operational needs
MAJOR FOCUS
Data

Governance of a non-existent layer
Architecture target/governance=
Common federated capabilities
Supports discovery, Open Standards
Common entities, Common data layer,
Common understanding/reuse of data

COMPREHENSIVE DATA PORTFOLIO PERSPECTIVE
Who, What, When, Where

DATA REPOSITORIES
Data Warehouses/Marts
Data Aggregation Systems
Data Consolidation Systems

ACCESS CONTROL

PRODUCTION SYSTEMS
Unique Component Transactional Systems

GOVERNANCE

FEW ENTERPRISE SYSTEMS
(5-15) SYSTEMS

(50-100) SYSTEMS

(600-800) SYSTEMS

Data Cleansing Data Standards
Authoritative data sources

Tagging people data rules

MANY COMPONENT SYSTEMS
Continued
Define a high level FRAMEWORK for governing and building big data

Define a target ARCHITECTURE that DHS Components can build to

Define requirements and STANDARDS to guide development efforts
BIG DATA AND DATA GOVERNANCE

- Apply DM Fundamentals for Governing and Building
- Ensure High Quality Content is Delivered to Big Data Stores
- Apply Good Governance to Manage Delivery Strategies

DHS Perspectives and Initiatives
APPLYING DM GOVERNANCE FUNDAMENTALS

ENSURING DATA MANAGEMENT’S ROLE WITH BIG DATA

BUSINESS CONSIDERATIONS
Who are the stewards/owners of the data sources and big data stores

IMPROVE UNDERSTANDABILITY
Are Data Models available to improve understandability
- Where does data come from and its lineage?
- What are the data definition standards?

INFORMATION SHARING RULES
Are Information Sharing Rules for Use, privacy and security well defined and available?

DATA QUALITY RULES
Are Data Quality standards defined and how will they be measured?

PRIVACY AND SECURITY
What is the privacy and security infrastructure for big data?

USE AND RETENTION
What are the rules for use and retention of the data by big data stores?

EXTRACTION AND ACCESS
What is the data tagging approach for extraction and accessing?

DHS Perspectives and Initiatives
OPERATIONAL CONSIDERATIONS FOR DATA SOURCES

- Is current or historical data needed, and how much?
- What are the extraction capabilities – services or file transfers?
- Who will do the data tagging?
- What data cleansing tools are being used?
- Are Use, access and role privilege information sharing rules available for each data store?

OPERATIONAL CONSIDERATIONS FOR BIG DATA STORES

- Are there common elements for connecting the data sources mapped to a common data model?
- How often is data needed for refresh or update?
- Are Use, access and role privilege rules defined?
GOVERNANCE OF DELIVERY STRATEGIES

DATA ANALYTICS

• What tools are being considered and used?
• What standards and best practices are available for improving data analytics

OPERATIONAL EXTRACTIONS FROM BIG DATA STORES

• Are there retention and disposal rules for operator extracts to their workspaces?
• What tools are being used and considered for managing operator workspaces?

DHS Perspectives and Initiatives
ENSURING HIGH QUALITY CONTENT

What are the quality and level of confidence requirements for Big Data?

Identifying what sources are authoritative and trusted is critical
- What are the authoritative data sources owners?
- Do the authoritative data sources meet data quality standards for Big Data?

How is data quality measured and reported? Dashboards, tracking reports, Profiling and Cleansing tool reports?

Has a risk and cost containment strategy for maintaining high quality Big Data been developed?
QUESTIONS & COMMENTS
CONTACT

MICHAEL SIMCOCK
DHS Chief Data Architect
OCIO/Enterprise Data Management Office
e: michael.simcock@hq.dhs.gov
p: 202.447.3740  |  m: 202.570.5320

PAUL REYNOLDS
Senior Information Architect
OCIO/ Information Sharing Environment Office
e: paul.reynolds@hq.dhs.gov
p: 202.617.5068