GEOINT Big Data: Implementing the Right Big Data Architecture

Todd G. Myers
Lead Global Compute Architect, NGA
2013 Joint GMU-AFCEA Symposium
• We must significantly improve how mission solutions are designed, developed, deployed and measured

• Cloud models need to support hosting real-time software engineering, service composition, and continuous integration processes

• Capabilities that are extensible, configurable mission services, APIs, and applications for on-demand, self-service analytics multi-tenant platform
The Problem…
Know The Earth... Show The Way... Understand the World

ANTICIPATE

Where something is... Why it's there... What may happen...

Target-Based  Activity-Based

Approved for Public Release  12-003
Volume

Terabytes to exabytes of existing data to process

New Sensor Capabilities Driving Need For New PED Solutions

- Sensor and Processing Data
- Spectral - more bands
- LIDAR
- Giga-pixel sensors
- 30 FPS WAMI
- Video SAR
- Higher quality Optics
- Source Metadata
- Records
- Transactions (Crowd Sourcing)
- Tables, Document, Chat
- Data from Humans
**Big Data Solutions: One Size ****DOES NOT** Fit All**

- Rapidly growing technology sector and an ecosystem that constantly is changing—need to find the right solution for your own unique circumstances

- NGA is developing a GEOINT Operating Framework which is a set of principles that directs architecture, standards, tradecraft, and enabling technology which will in turn guide how our workforce, content, and systems interact to produce GEOINT products and services.
Information about the data…

Information about the data,

metadata

about the information about the data,

entity extraction

about the information about the data for knowledge to intelligence.

contextual resolution
### Information about the data

**GEOINT types of data are vast**

<table>
<thead>
<tr>
<th>Relational Data</th>
<th>Phenomenologies</th>
<th>Platforms</th>
<th>Providers</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Systems</td>
<td>Geophysics</td>
<td>Motion/Video</td>
<td>National</td>
</tr>
<tr>
<td>Content Management</td>
<td>Hyerspectral</td>
<td>Panchromatic</td>
<td>DoD</td>
</tr>
<tr>
<td>Email</td>
<td>Infrared</td>
<td>Polarimetric</td>
<td>Local/State</td>
</tr>
<tr>
<td>CRM</td>
<td>LIDAR</td>
<td>Radar</td>
<td>Commercial</td>
</tr>
<tr>
<td>Supply Chain</td>
<td></td>
<td>Airborne</td>
<td>Foreign/International</td>
</tr>
<tr>
<td>ERP</td>
<td></td>
<td>Spaceborne</td>
<td>Open Source</td>
</tr>
<tr>
<td>RSS Feeds</td>
<td></td>
<td>Handheld</td>
<td></td>
</tr>
<tr>
<td>Cloud</td>
<td></td>
<td>Surface</td>
<td></td>
</tr>
<tr>
<td>Custom Sources</td>
<td></td>
<td>Subsurface</td>
<td></td>
</tr>
<tr>
<td>External Sources</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Foundations**
- Elevation
- Features
- Gravity
- Controlled Imagery
- GPS
- Soils
- Bathymetry

**NGA**
- COCOMS
- Commercial
- Open Source

**Requests**
- Service/Intel Centers
- Intel Agencies
- International

**Services**
- Request
- Tasking
- Collection
- Discovery
- Retrieval
- Dissemination
- Processing
- Analytic
- Archive

**User Problem Solver**
- Detect
- Classify
- Characterize
- Identify
- Understand

**Structured Data**
- Transactional
- Time phased data

**Semi &/or Unstructured Data**
- Text Report
- Emails
- Presentations
- Videos

**Patterns**
- Point of interest over time

Bruce Brown, Big Data Architect
brownb@us.ibm.com

Imagery Copyright 2012 Digital Globe, GEO Eye, PA Dept. of Conservation and Natural Resources – PAMAP/USGS

Approved for Public Release 10-231 & 12-446
about the information about the data

Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations

- Metadata
- Standards
- Tagging
- Ontologies
- Discoverability
- Governance
- Accuracy
- Precision
- Data Quality
- Data Retention
- Analytic Confidence/Rating
- Calibration
Visualization

Activity Based Intelligence (ABI) will form “the intellectual underpinning for how we conduct intelligence in the future”

Activity and transactions are the fundamental building blocks

Extracted activity contains vital intelligence information

ABI optimizes exploitation of large scale data

Goals: Discover, find and characterize activity, patterns of life, networks and anomalies
about the information about the data for knowledge to intelligence

Activity-Based Intelligence in Action

Sources
Automated processing of raw source data into vector features and activities.

Processing

Foundation

Observables

Auto Tip-off

Data Needs

Intelligence Questions

Scenario
Activities/Events
Actors
Equipment
Locale

Analysis
Non-linear analysis to identify linked activity of interest, make judgments, and task collection.

Knowledge Capture

JUDGMENT

ANSWERS

EVIDENCE

Understanding

Activity

Networks

Intelligence in Action Only!

Vector information about activities, events, entities, and locations on-line/on-demand.

Normalized Data Services

Approved for Public Release 12-446

Imagery Copyright 2012 Digital Globe, GEO Eye, PA Dept. of Conservation and Natural Resources – PAMAP/USGS
Questions?