GEOINT Big Data:
Implementing the Right Big Data Architecture

Todd G. Myers
Lead Global Compute Architect

Big Data Symposium
Washington, DC • September 25, 2013
NGA Has a Vital Mission

The National Geospatial-Intelligence Agency (NGA) provides timely, relevant, and accurate geospatial intelligence – or GEOINT – in support of national security objectives.

NGA’s Mission Set

- Strategic Intelligence
- Warfighter Support
- Counterterrorism
- Counterproliferation
- Cyber Warfare
- Homeland Security
- Safety of Navigation
- Humanitarian Relief
- Foundation Data

“NGA has grown into a critical link in America’s intelligence apparatus and made a decisive difference to our national security and recent war effort.”

– Robert M. Gates, Former Secretary of Defense
NGA Utilizes a *Wide Range* of Data

**Remotely Sensed Data**
- Panchromatic
- Infrared
- Multispectral
- Hyperspectral
- Radar

**Foundation Data**
- Physical Geography
- Maritime Safety of Navigation
- Aeronautical Safety of Navigation
- Gravitational Model
- Culture Data

**Land**
- Feature Data
- Vegetation
- Terrain
- Maps
- Controlled Imagery

**Sea**
- Safety of Navigation
- Maritime Information
- Maritime Surveys
- Charts
- Notice to Mariners

**Air**
- Safety of Navigation
- Aeronautical Information
- Aeronautical Surveys
- FLIPS
- Charts
- Vertical Obstructions

**Science**
- GPS
- Photography
- Gravity
- Coordinate Systems

**Research**
- Human Geography
- Geographic Names
- Open Source
- Maritime and Land Boundaries

Approved for Public Release – NGA Case #13-153
• 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…

THE INTERNET in 2015
Is the dawn of the Zettabyte era
But how much data are we talking about?

1000 megarabytes = 1GB
1000 gigabytes = 1TB
1000 terabytes = 1PB
1000 petabytes = 1EB
1000 exabytes = 1 ZB

How much is an exabyte really?
1 exabyte amounts to 36,000 years of HD-TV video, or the equivalent of streaming the entire Netflix catalog 2.877 times.

WHAT IS A PETABYTE?
To understand a petabyte we must first understand a gigabyte.

1 gigabyte = 7 minutes of HD-TV video
2 gigabytes = 20 yards of books on a shelf
4.7 gigabytes = size of a standard DVD-R

A PETABYTE is a lot of data

1 petabyte = 20 million 4-drawer filing cabinets filled with text
1.5 petabytes = size of the 10 billion photos on Facebook
15+ petabytes = internet users data backs up on Mozy.com
20 petabytes = the amount of data per processed by Google Day
20 petabytes = total hard drive space manufactured in 1995
50 petabytes = the entire written works of Shakespeare, the entire run of the New York Times, or 124 years of recorded history.

By 2015, nearly 3 billion people will be online, pushing the data created and shared to nearly 8 zettabytes.

IT WOULD TAKE OVER 5 YEARS TO WATCH THE AMOUNT OF VIDEO THAT WILL CROSS GLOBAL NETWORKS EVERY SECOND in 2015

How will internet traffic grow?

What type of video?

How will people be getting their video?

Most internet traffic in 2015 will be VIDEO

What does internet traffic look like right now?

http://mrpalsmy.wordpress.com/category/btechno/
Approved for Public Release – NGA Case #13-453
NGA is at the *Forefront* of ABI

**Activity-Based Intelligence (ABI) is advancing and evolving GEOINT**

- Activity-Based Intelligence is a set of methods or workflows that enable better intelligence through knowledge discovery and capture.
- ABI focuses on capturing activities as they occur and – based on understanding of patterns of life – analyzing those activities to determine normal from abnormal, to determine relationships, and to discover networks.
- This would include activities associated with culture, religion, economic, societal, etc.

“We are being challenged to think in terms of activity-based GEOINT rather than target-based GEOINT and to explain not only where something is happening, but also why.”

– Letitia A. Long, Director, NGA
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

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

Phenomenologies
- Geophysics
- Hyperspectral
- Infrared
- LIDAR
- Motion/Video
- Panchromatic
- Polarimetric
- Radar

Platforms
- Airborne
- Spaceborne
- Handheld
- Surface
- Subsurface

Providers
- National
- DoD
- Local/State
- Commercial
- Foreign/International
- Open Source

Foundation
- Elevation
- Features
- Gravity
- Controlled Imagery
- GPS
- Soils
- Bathymetry

Providers
- NGA
- COCOMS
- Commercial
- Open Source
- 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

VARIETY
Data in many forms

Activity

Networks

Associations

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

Intelligence in Action Only!
Understanding

Networks
Activity

EVIDENCE

ANSWERS

JUDGMENT

Tasking

Knowledge Capture

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

Approved for Public Release 12-446
Bottom-Line

Getting GEOINT to the Tactical Edge
Questions?
NATIONAL GEOSPATIAL-INTELLIGENCE AGENCY
Know the Earth... Show the Way... Understand the World

nga.mil | @nga_geoint | facebook.com/natlgeointagency