Geospatial Cloud Analytics:
The Confluence of Commercial Space and the Computing Revolution

Dolores Shaffer, Science and Technology Associates

for

Dr. Joseph B. Evans, Program Manger, DARPA/STO

Chesapeake Large-Scale Analytics Conference

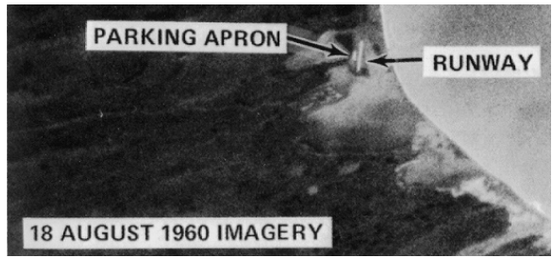
30 October 2018

Outline

- Trends in geospatial data and computing
- Geospatial Cloud Analytics: origins and objectives
- Scalable Data Platforms
- Analytics

Geospatial awareness – everywhere, anytime

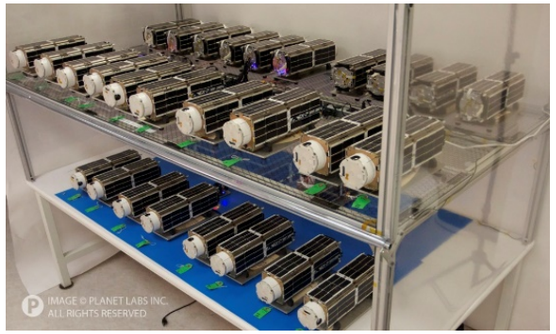
Early Space Age



Sensing

Today

Revisit



Computing

Lack of global data

- Government satellites – high resolution, limited coverage, stovepiped operations
- Traditional commercial satellite constellations – limited coverage, and fragmented marketplace
- No comprehensive commercial repository of multimodal data exists

Earth: 510 million km^2
Earth's Landmass: 148 million km^2

Imagery Source	Resolution	Revisit Rate	Coverage per Day
Landsat NIR/SWIR	30m	Weekly	13 million km^2
Sentinel 1	20m	Weekly	
Sentinel 2	10m	Weekly	11.6 million km^2
WorldView 2	1.85m	3.7 days	1 million km^2
WorldView 3	1.24m	4.5 days	0.68 million km^2

NIR/SWIR: Near-Infrared/Short Wave Infrared

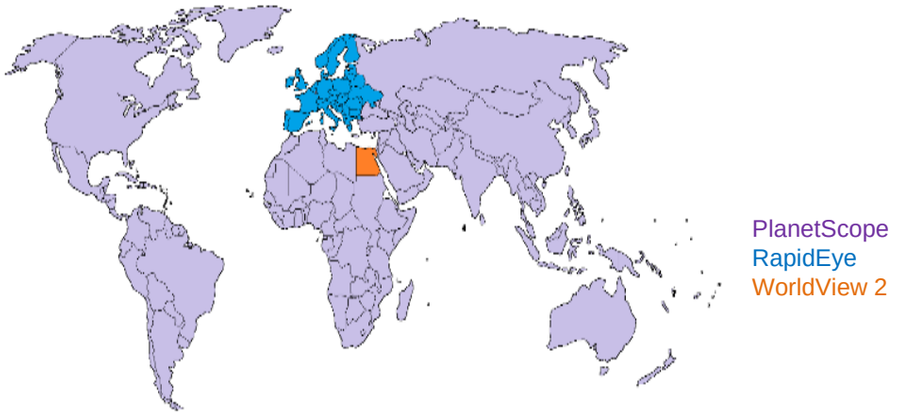
Filling the gap: New commercial space industry – small satellites

- New, large constellations offer extensive coverage with a high revisit rate

Earth: 510 million km^2
Earth's Landmass: 148 million km^2

Imagery Source	Resolution	Revisit Rate	Coverage per Day
RapidEye	5m	5.5 days	6 million km^2
SkySat	0.9m	Daily	21 million km^2
PlanetScope	3m	Daily	150 million km^2

Visualizing the Daily Coverage Footprints

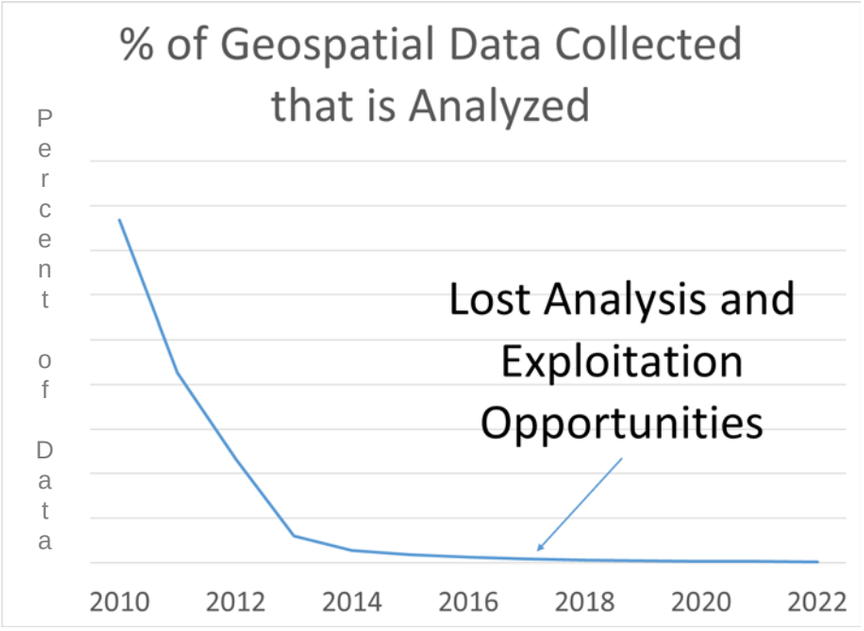


- Multiple modalities – optical, SAR, RF
- Really big data
  - PlanetLabs' global coverage with daily revisit = 9 PB/year
  - Digitalglobe collects 100 TB/day = 36 PB/year, various bands and resolutions

Current approaches limit exploitation of new sensor constellations

- Inability to scale analytic processing and techniques
  - Data analysis dominated by labor intensive, manual techniques
  - Does not scale to tsunami of data from new satellite constellations

- Constraints on analytics services for DoD users
  - Typical path is through defense intelligence community

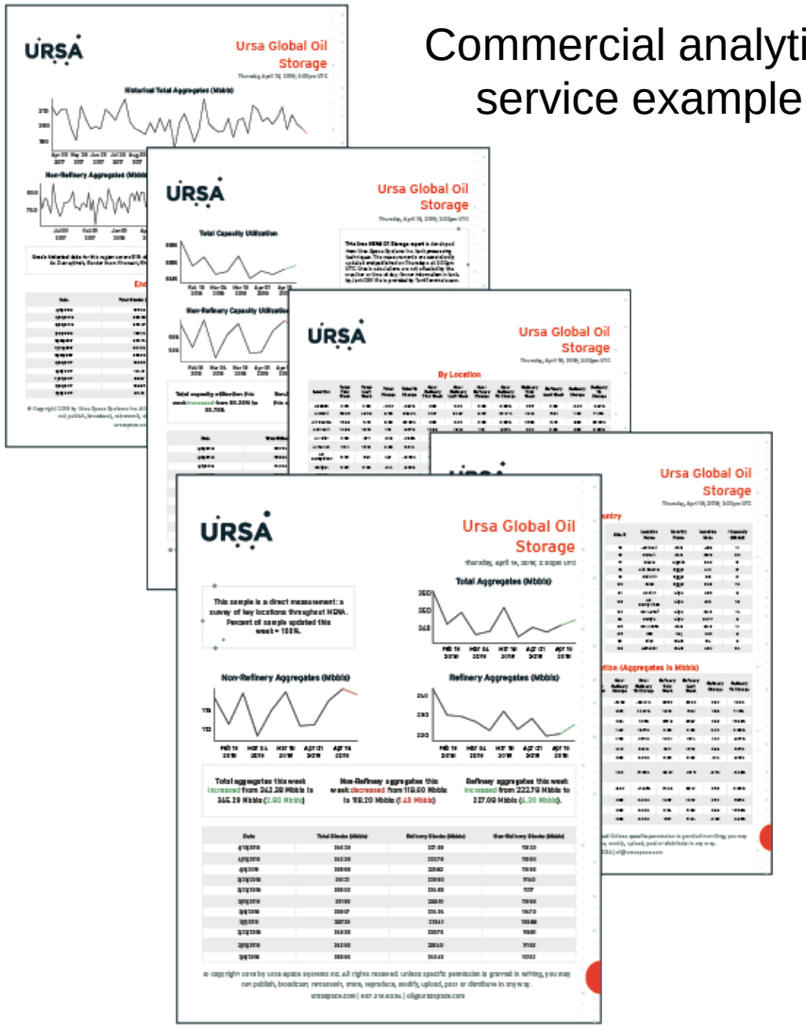


There are new opportunities for exploitation

- Vast amount of data from new commercial satellite constellations enables exploitation of geospatial analysis as a big data problem
  - Research has demonstrated the ability to use disparate data sources such as Planet imagery, OpenStreetMaps, and Wikipedia to observe maritime behavior

- Confluence of technologies – combine extensive coverage and high revisit rates with cloud, big data, machine learning approaches

- Today: days-to-weeks timelines limited by data collection and processing
- Tomorrow: Global, continuous automated multi-source change detection



Commercial analytics service example

Can we provide analysis services rather than pixels for DoD warfighters?

Geospatial Cloud Analytics (GCA)

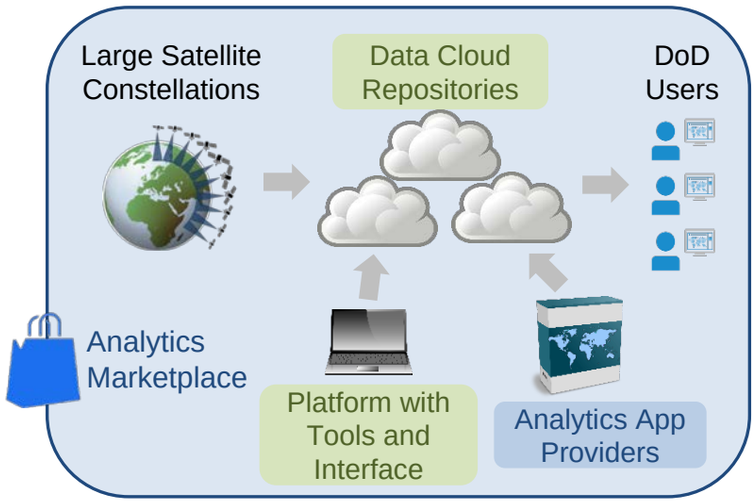
GCA will demonstrate rapid global analytics

- Objective – Develop technology to access & analyze global scale commercial geospatial data and pilot an analytics-as-a-service business model

Current approach cannot satisfy analytics demand

- DoD (e.g., COCOM) cannot monitor militarily relevant events/changes on a global, near real time scale
- Global scale analytics need multimodal data
- Global-level I&W missions / analytics do not exist

GCA approach

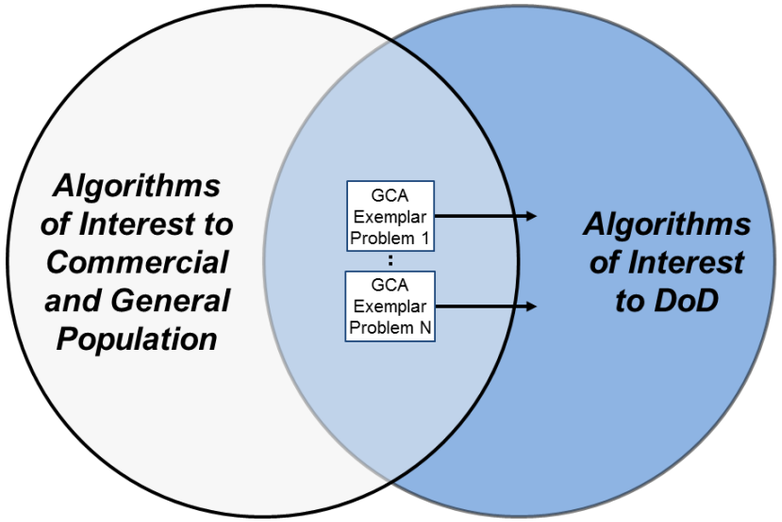


- Vision: Let experts focus on analyzing - versus gathering and curating - data from individual sources
- Create cloud-based multi-source/multi-modal data repository
- Demonstrate ability for analytics to scale globally
- Create analytics marketplace to ease DoD use

I&W: Indicators and Warnings COCOM: Combatant Command

Commercial geospatial data and analytics can be applied to DoD problems

- Many analytics of value to non-DoD applications are directly or near-directly applicable to DoD analytics needs
  - I&W of major events, threat activity / trends, etc.
  - Find and track fleeting high-value targets
  - Discover hidden and unknown sites, and understand purpose of enigma sites
  - Discover and provide attribution for gray zone



Processing and algorithmic scalability by coupling commercial scale and novel analytics

A new approach – global cloud repository, scalable analytics, competitive marketplace

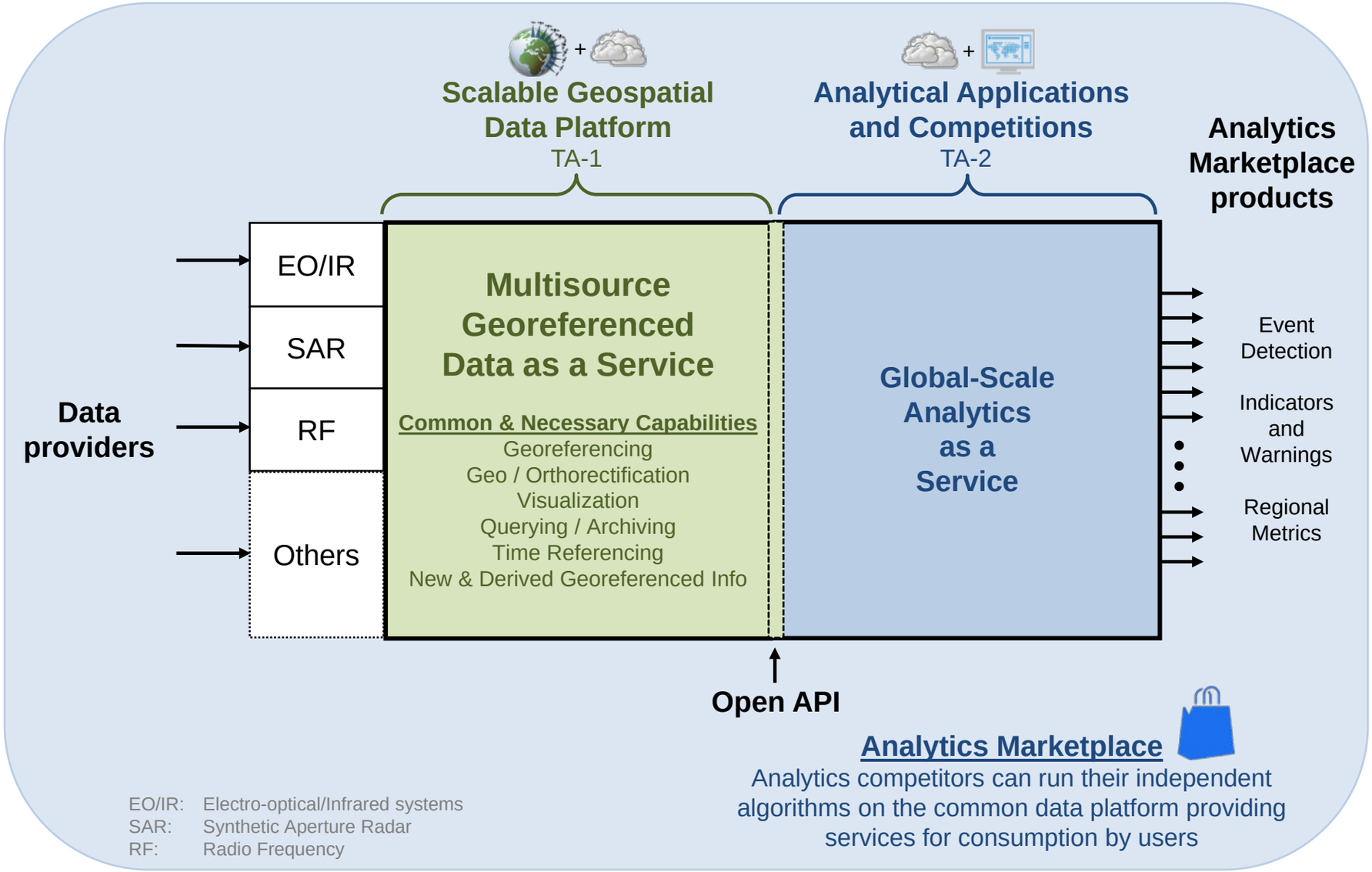
- Cloud-based multi-source/multi-modal data repository
  - Virtually centralized access to physically distributed data
  - Automated tools to curate data & metadata

- Scalable analytics
  - Global coverage, multi-modal processing
  - Co-use or leveraged analytics techniques

- Competitive analytics marketplace
  - Explore business models that encourage participation of data, analytics platform, and 3rd party app providers
  - For example, “app store” analytics platforms that create a sustainable ecosystem

Today: Days-to-weeks timelines limited by data collection and processing
Tomorrow: Global, continuous automated multi-source change detection

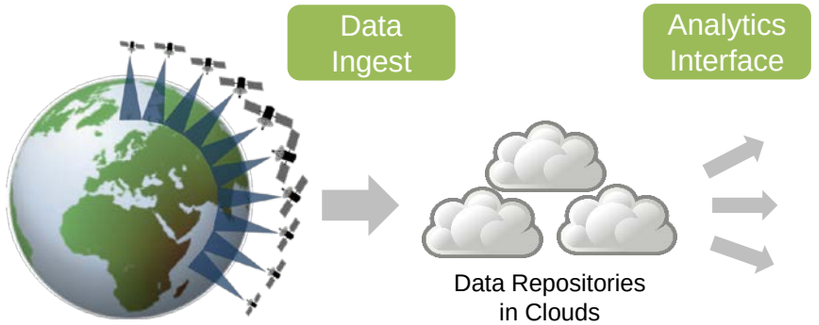
Program structure



TA-1: Scalable geospatial data platform

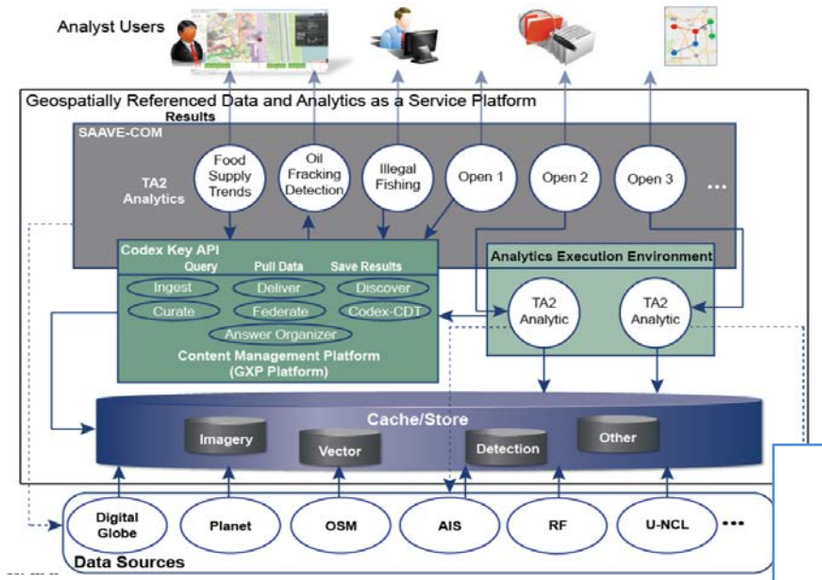
- TA-1 enables TA-2 to focus on analysis
  - Gather and clean data
  - At least 10 sources
    - EO/IR
    - SAR
    - RF
    - Other
  - Fast query capability

- TA-1 performers
  - BAE Systems
  - Descartes Labs
  - Digitalglobe

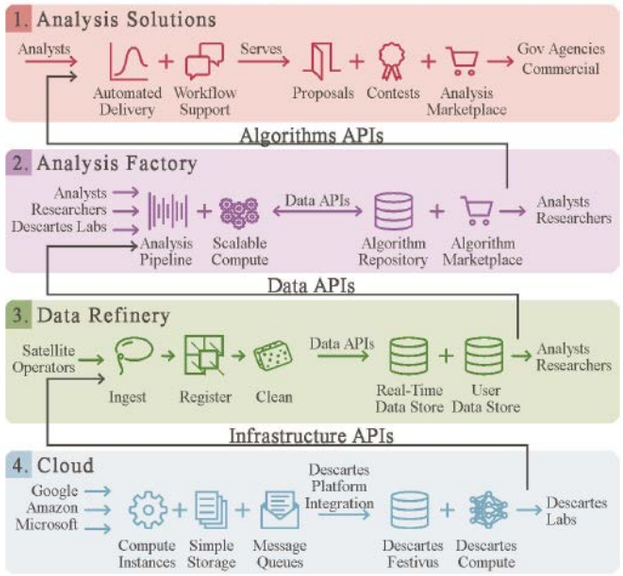


Scalable, cloud-based platform for easy ingestion and use of satellite imagery

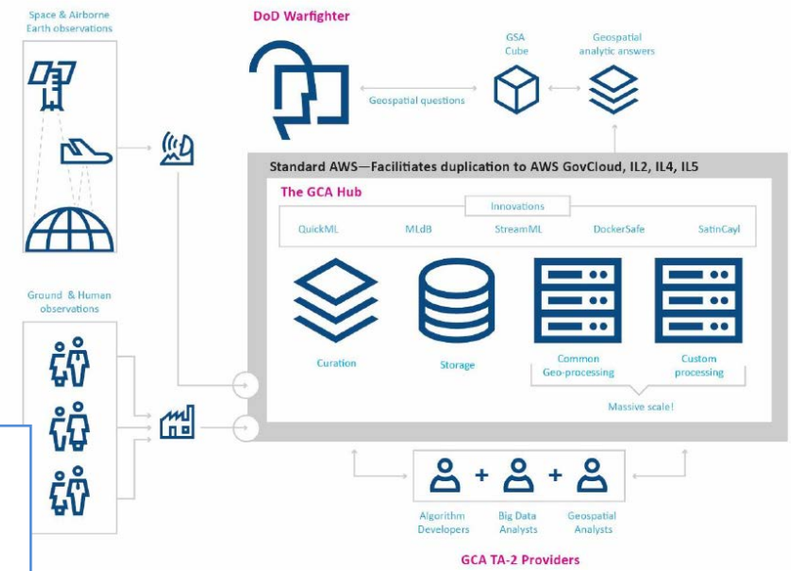
GCA performers
TA – 1: Scalable geospatial data platform



BAE



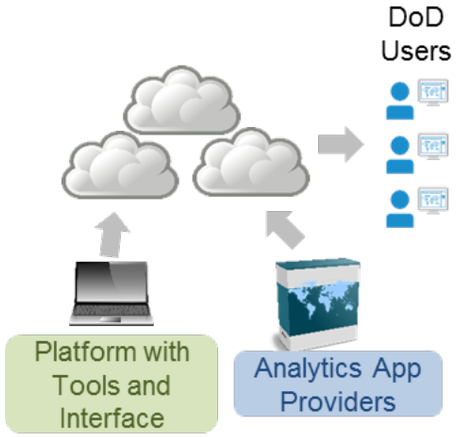
Descartes Labs



DigitalGlobe

TA-2: Analytical applications and demonstrations

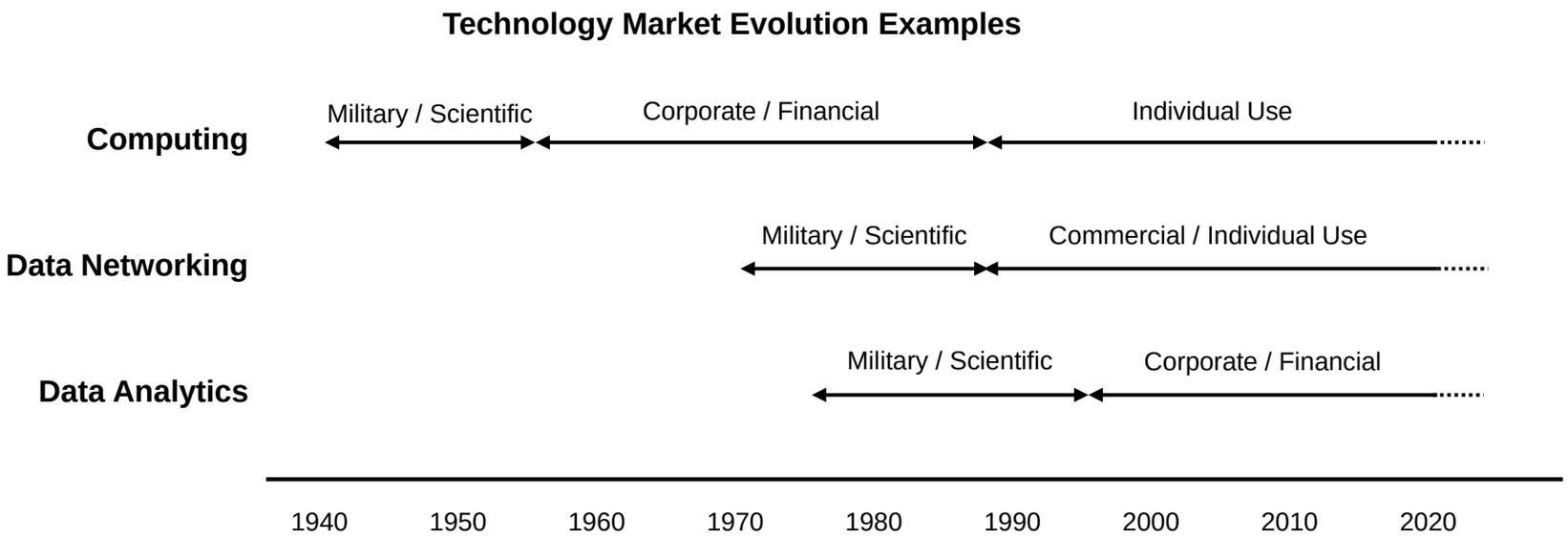
- Challenges
  - Scaling applications for global coverage
  - Exploiting a mixture of sensor modalities for analytics
- Approach: challenge problem areas
  - Four problems (strategic, tactical, operational, and open)
  - Demonstrations designed to validate scalability and analytic utility
- Solutions
  - Remote sensing techniques
  - Machine learning
  - Hybrid solutions
  - Multiple data sources to reduce processing and improve accuracy



GCA TA-2 performers

	Strategic	Operational	Tactical	Open	Notes
Textron	X				Global food forecasting platform
Descartes Labs	X	X			Strategic: Food security in selected locations Operational: Automated deep learning broad area search for fracking activity
Lockheed Martin		X			Convolutional neural networks for classification
Raytheon			X		Pattern of life characterization
BAE Systems			X		Incorporation of behavioral data
SRI				X	Trafficability analysis
DZYNE				X	Humanitarian assistance and disaster recovery assessment (HA/DR)
VSI				X	Global digital elevation model (DEM)
Draper				X	Economic instability reports
Analytic Strategies				X	Global protest detection using mobile data

GCA analytics marketplace



- Service Level Agreements (SLAs) for analytics services
- Government purchases services, rather than software or labor

Objective: Lower barriers to entry to providing DoD analytics
→ More non-traditional performers and capabilities
→ Look at more problems
→ Stay on the cutting edge

GCA Phase 3

- Phase 3 Overview
  - This BAA solicits only for Phases 1 and 2
  - Phase 3 will be a separate solicitation and will focus on creating an analytics-as-a-service marketplace

- Business model
  - Proposals should include a business model to serve as the basis for Phase 3
  - This will be further refined to reflect lessons learned during Phases 1 and 2
  - Example
    - The Government may ask for the answer to a specific question about crop growth trends over the past 10 years in a particular region, then competitively procure an answer to the question through the GCA marketplace.

TA-2: Food security analytics

- Problem area: Predicting food supply trends for a crop in a region of the world
- Timeline: Weeks to months
- Why we care: precursor to civil unrest
- Conventional approach relies on government reporting
  - Timeliness
  - Accuracy

- Performers
  - Textron – subcontractor Geosys, a Land O'Lakes company
  - Descartes Labs (example follows)

Food security example: Descartes Labs approach

Technical Approach

Segment

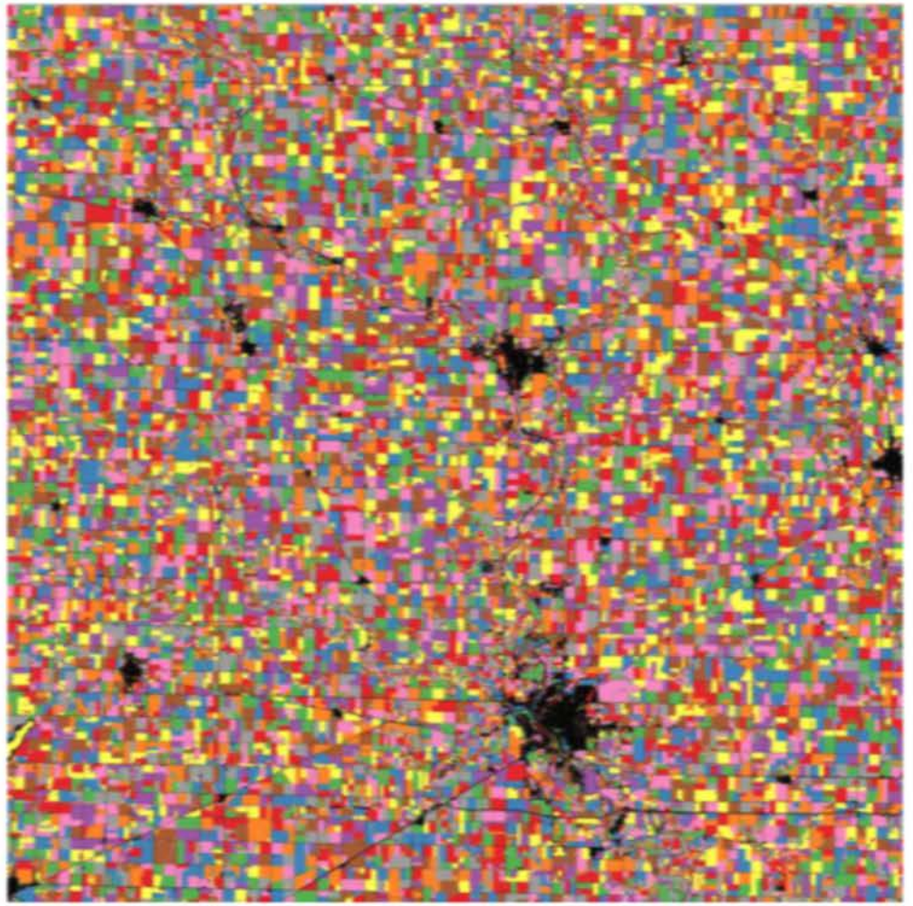
- Detect functional field boundaries
- Identify persistent edges across a multisensor deep temporal stack

Classify

- Pixel-level classification of Wheat vs Non-Wheat, per season.
- Useful both for acreage and as a mask for health metrics.

Monitor

- Develop crop health metrics
- Relative crop conditions derived from NDVI
- Ideally, measure absolute conditions, e.g. yield (kg/ha) and production (kg)



Food security example: Descartes Labs approach

DESCARTES LABS

Technical Approach

Segment

- Detect functional field boundaries
- Identify persistent edges across a multi-sensor deep temporal stack

Classify

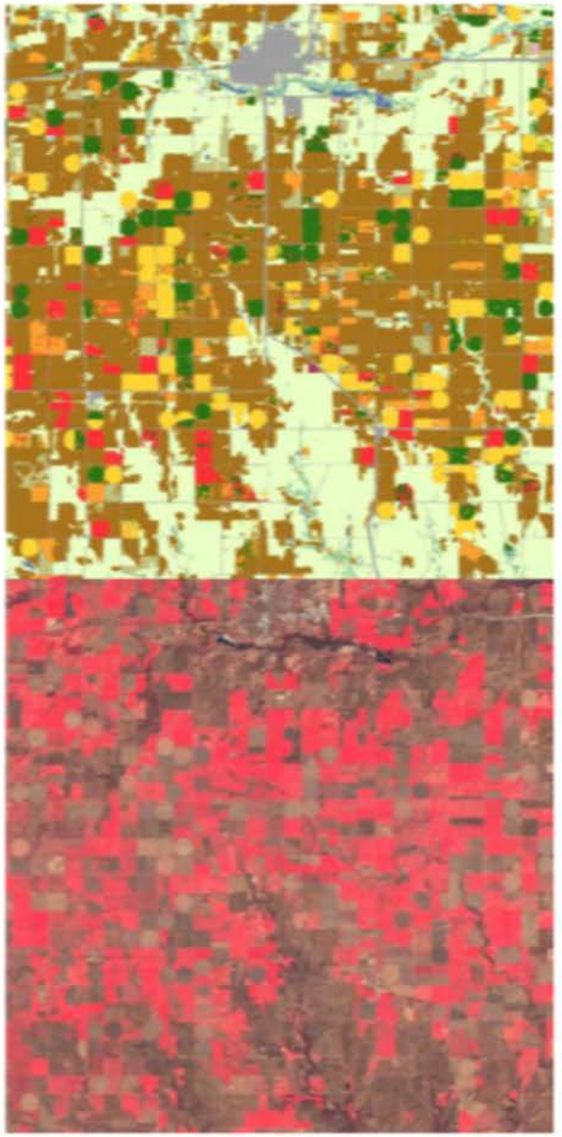
- Pixel-level and field-level classification of Wheat vs Non-Wheat, per season.
- Useful both for acreage and as a mask for health metrics.

Monitor

- Develop crop health metrics
- Relative crop conditions derived from NDVI
- Ideally, measure absolute conditions, e.g. yield (kg/ha) and production (kg)

Cropland Data Layer

Sentinel-2 NIR



Food security example: Descartes Labs approach

DESCARTES LABS

Technical Approach

Segment

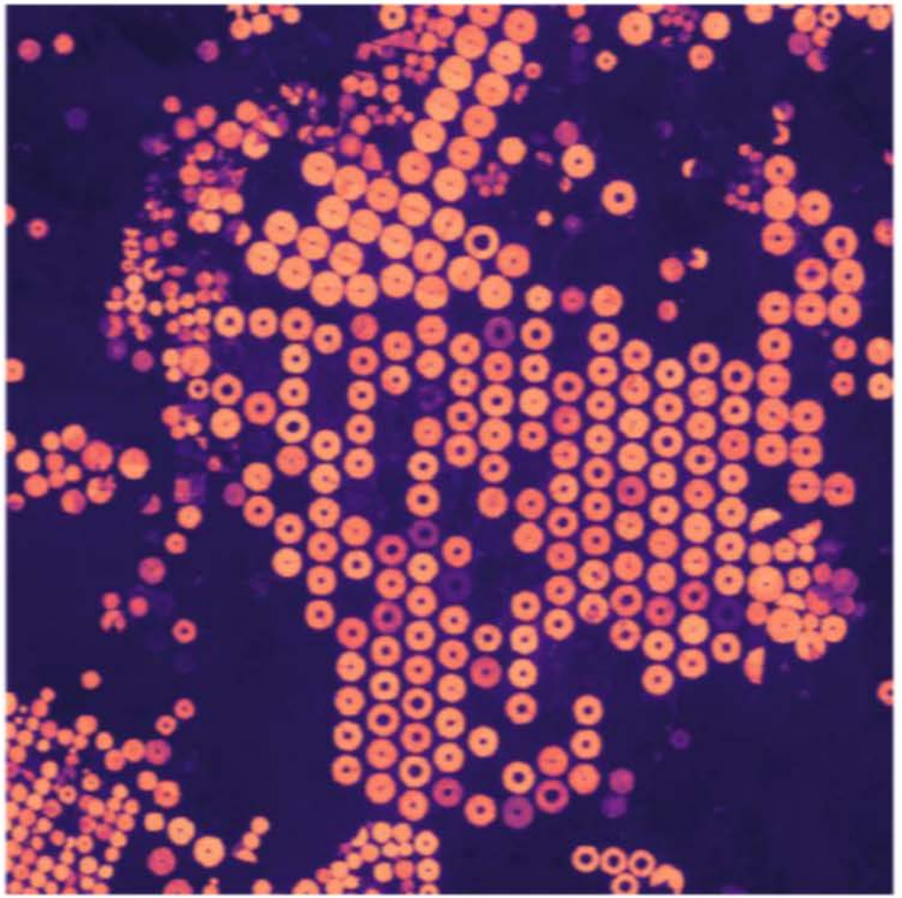
- Detect functional field boundaries
- Identify persistent edges across a multi-sensor deep temporal stack

Classify

- Pixel-level classification of Wheat vs Not-Wheat, per season.
- Useful both for acreage and as a mask for health metrics.

Monitor

- Develop crop health metrics
- Relative crop conditions derived from NDVI
- Ideally, measure absolute conditions, e.g. yield (kg/ha) and production (kg)



TA-2: Fracking construction detection analytics

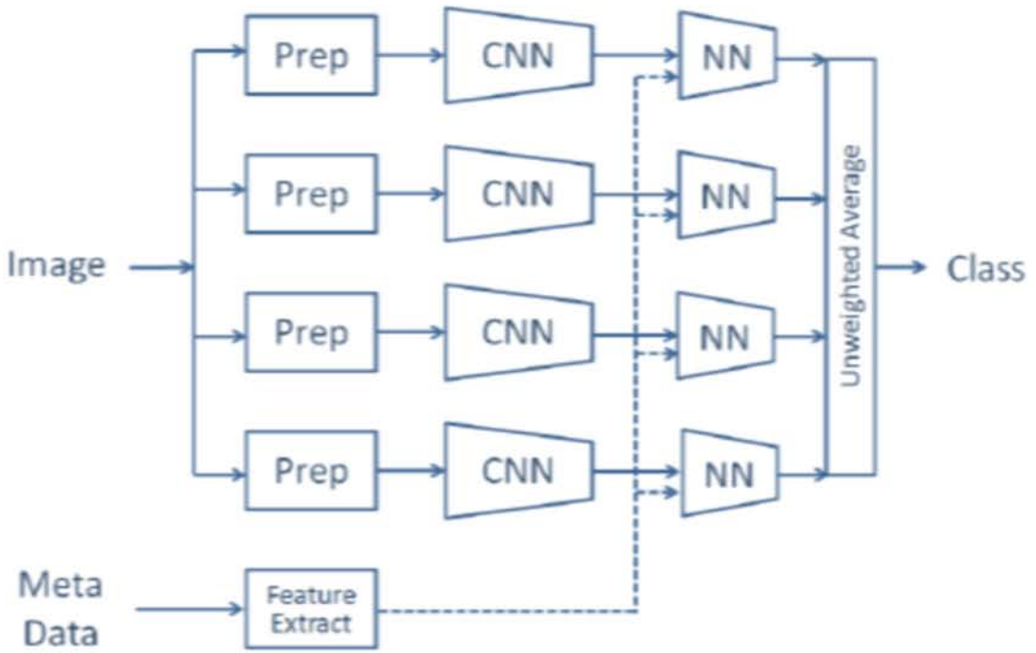
- Problem area: Locating the construction of oil fracking sites
- Timeline: Days to weeks
- Why we care: oil supply monitoring
- Constructed in stages
  - Monitoring
- Conventional approach is reviewing permit information, visiting sites

- Performers
  - Descartes Labs – subcontractor Draper
  - Lockheed Martin (example follows)

Fracking detection example: Lockheed Martin approach

OUR SOLUTION: CNN ENSEMBLE WITH METADATA

- We developed an ensemble of deep convolutional neural networks (CNNs).
- We integrated satellite image metadata in the system:
  - Improved accuracy by 3.5%
- We coded in Python and used open source software for deep learning:
  - Keras
  - TensorFlow
- We classified image chips into 62 object and facility classes
  - Runways, ports, shipyards, towers, oil facilities, tunnel openings, etc



M. Pritt and G. Chern, "Satellite Image Classification with Deep Learning," *Proceedings of IEEE AIPR Workshop*, 12 Oct 2017

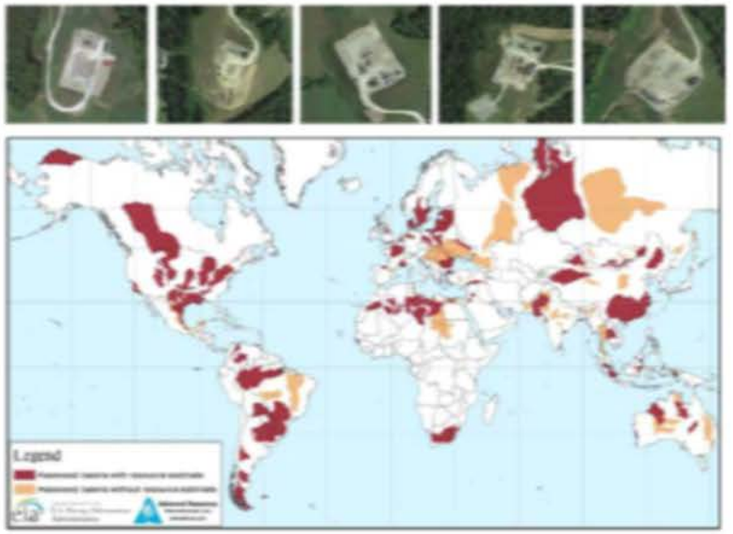
LOCKHEED MARTIN

Fracking detection example: Lockheed Martin approach

DARPA GCA: FRACKING DETECTION

- We will extend our algorithms to the problem of detecting oil and gas fracking wells.
- Fracking wells are large structures with distinctive features and temporal changes:
  - Pad construction: land clearing, grading
  - Drilling: heavy equipment and drilling rig
  - Fracking: more heavy equipment and pond
  - Production: well pad shrinks

2008 2010 2012 2015



Goal: Regional detection and monitoring

LOCKHEED MARTIN

TA-2: Illegal, Unreported, & Unregulated (IUU) fishing detection

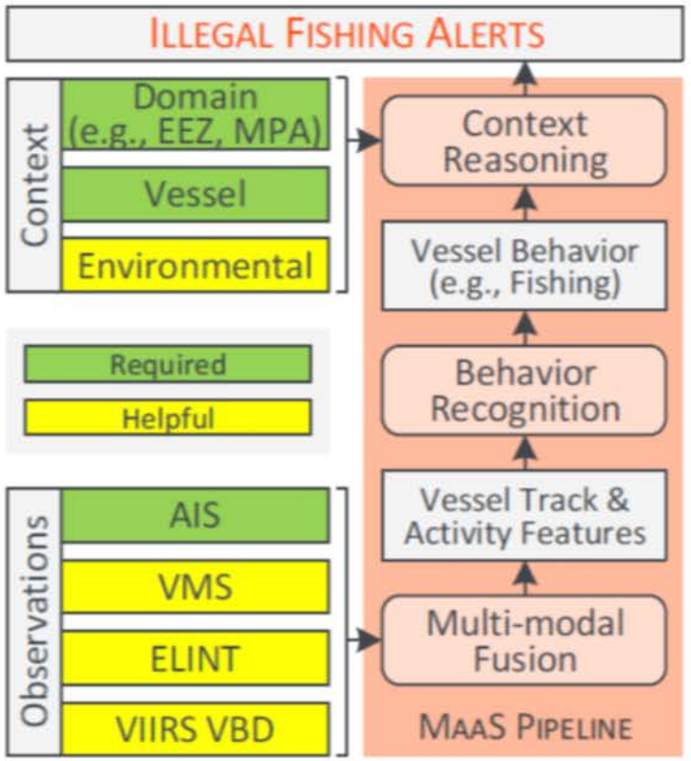
- Problem area: Maritime change detection and illegal fishing
- Timeline: Minutes to hours
- Why we care:
  - 3+ billion people rely on seafood as their primary source of protein
  - 90% of the world's fish stocks are either overfished or fully exploited
  - $23+ billion worth of seafood is stolen from the seas annually
  - Contributes to civil unrest
  - Similar to detecting other ocean-based illegal activities
- Conventional approach: patrol boats

- Performers
  - Raytheon
  - BAE Systems (example follows)

IUU example: BAE Systems approach

Multi-INT Analytics for Pattern Learning & Exploitation (MAPLE) as a Service (MaaS)

- MaaS monitors maritime activity in three steps:
  - Multi-modal Fusion (MMF) learns discriminative features via compact joint representations over multiple sources
  - Behavior Recognition (BR) learns novel classes from few examples in category hierarchies
  - Context Reasoning (CR) identifies when vessel behaviors & domain conditions are consistent with illegal fishing
- Maximal exploitation of data improves results
  - MMF treats data sources jointly
  - BR model complexity is data-driven
  - CR performs probabilistic inference over domain-specific data



BAE SYSTEMS

TA-2: Open category analytics

Who	What	Why
Analytic Strategies	Protest Detection	Detect unrest
Draper	Instability Index	Detect unrest
DZYNE	Trafficability	Humanitarian Assistance / Disaster Recovery
SRI	Trafficability	Humanitarian Assistance / Disaster Recovery
VSI	3D DEM	Disaster assessment

Protest detection: Analytic Strategies approach

Overview

- Open Call Competition: *Global Protest Detection Algorithm* will use device data that can emanate from over half of the world's population
  - Uses commercial high-quality location data previously only available via classified signals intelligence (SIGINT)
  - Does not rely on billion-dollar satellites, network taps, human assets, or other means of collection
  - Provides access to a large number of devices running ad-enabled apps, growing in size with IoTs
  - Publicly-available information
  - Anonymous in raw form
  - Platform, operating system, network, and cell independent
  - SIM independent - an ad ID to a physical device remains the same unless it is reset
- Data works day or night in cloudy or sunny conditions; ubiquitous
- We seek better use cases for this data source, versus "lone wolf"
- *Data Scales globally, and it is possible to provide NRT alerts*

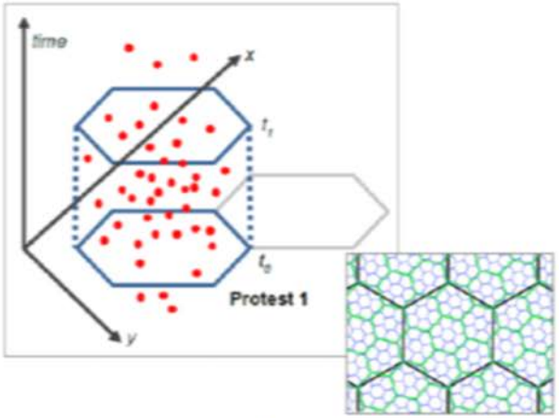
CULMEN INTERNATIONAL, LLC

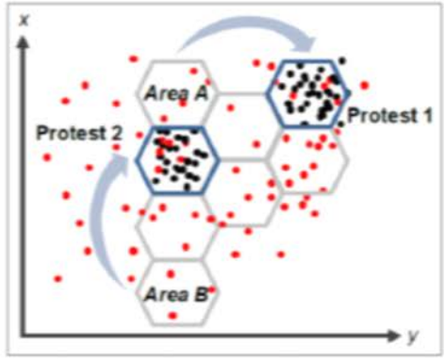
Facilitating Collaboration and Enhancing Security Worldwide.

1

Protest detection: Analytic Strategies approach

Approaches





- Recognize salients in space & time over the norm in high-risk areas leveraging a localized DGG
- Train model to recognize high-risk areas and repeat devices on prior observed protests

- Recognize aggregations from areas that conform to similar areas of origination
  (here Area A and Area B are statistically very similar as well as destination locations)
- Train model to recognize movement patterns from different device clusters but having similar contexts or profiles

- Geographic Diversification
- Performance Segmentation
  Seek utility in size, type, vendor, and time
- Increment Effectiveness
  Focus on gains in model effectiveness

CULMEN INTERNATIONAL, LLC

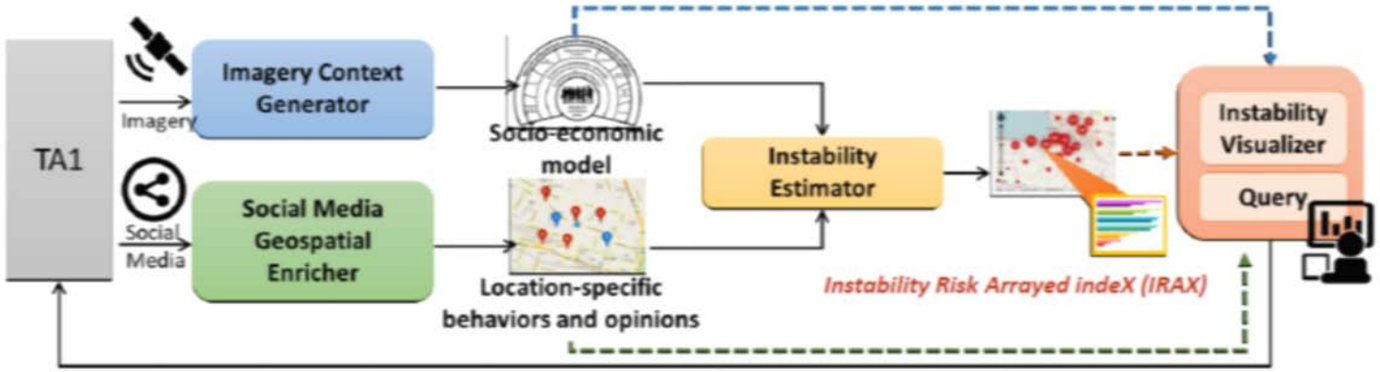
Facilitating Collaboration and Enhancing Security Worldwide.

2

Instability index: Draper approach

Global Anticipation of Instability through Novel Economic Reasoning (GAINER)

- ***TA2 Open Challenge Problem: Generate localized, reliable, timely instability estimates to support military leaders at all echelons***



- Fuse stable contextual information with transient social indicators
  - **Stable:** socio-economic indicators from commercial imagery, linking physical structures and societal features
  - **Transient:** location-specific PMESII information extracted from social media
- Cloud-based operation at scale – near real time

DRAPER

2

HA/DR: DZYNE Technologies approach

Road closure due to washed out road

Road closure due to down powerlines

Road closure due to elevated water level

DZYNE Technologies HA/DR Open Competition

- Road traversability geospatial analytics
  - Identify road blockage
    - Traffic congestion, debris, washed out road
  - Characterize level of blockage
    - Traversable by various class of transport
  - Define feasible routes to critical infrastructure
    - Power station, water treatment, hospital, police/fire station, government building
- Real-time analysis of affected area
  - Live updates of road conditions
  - Dynamically analyze feasible routes
- Image resolution requirement
  - Temporal resolution - hours to weeks
  - Spatial resolution - 30 cm - 1 m
  - Elevation angle - 60 - 90 degrees

Geospatial situational awareness for disaster management

HA/DR: DZYNE Technologies approach

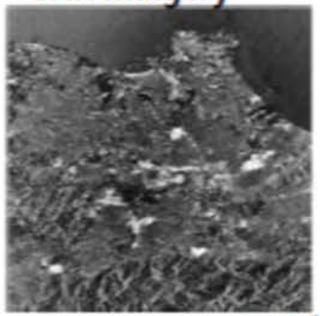
DZYNE

Deep Learning Innovation

Satellite Imagery

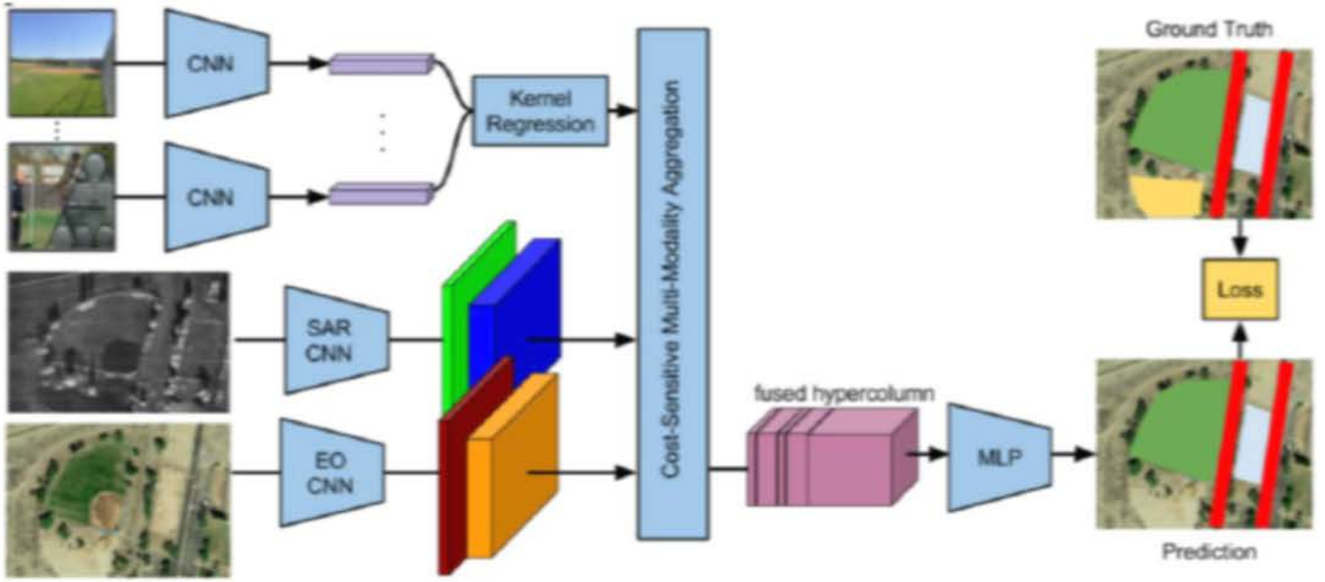
Ground level Imagery

SAR Imagery



Social Media

Cost-Sensitive Multi-Model Aggregation scalable and adaptable to new data sources.

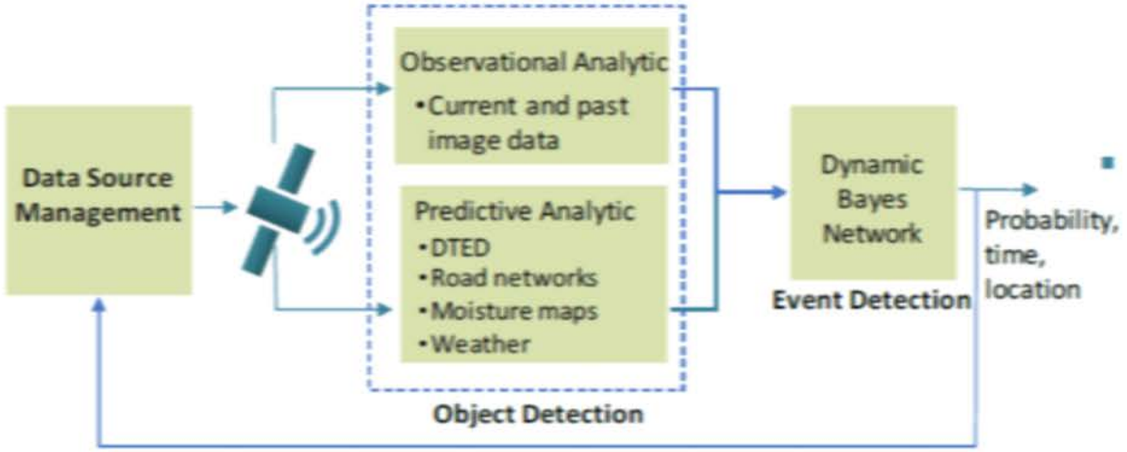


Auto-Tagging of building and region instances increases inferences of routable network. Pixel-wised detected roads are converted into a vector model for greater flexibility of traversability analysis.

Trafficability: SRI approach

SURPAS: SUpply Route Predictive AnalyticS

- Supports optimized route planning by detecting and predicting objects and events
  - Flash floods, landslides, obstructions
  - Non-trafficable or poor quality roads and local bypass shoulders
- Accurate predictions of terrain dynamics with
  - Fully distributed, processed based hydrology model
  - Precision soil moisture mapping model
  - Dynamic Bayesian Network (DBN) event risk probability model

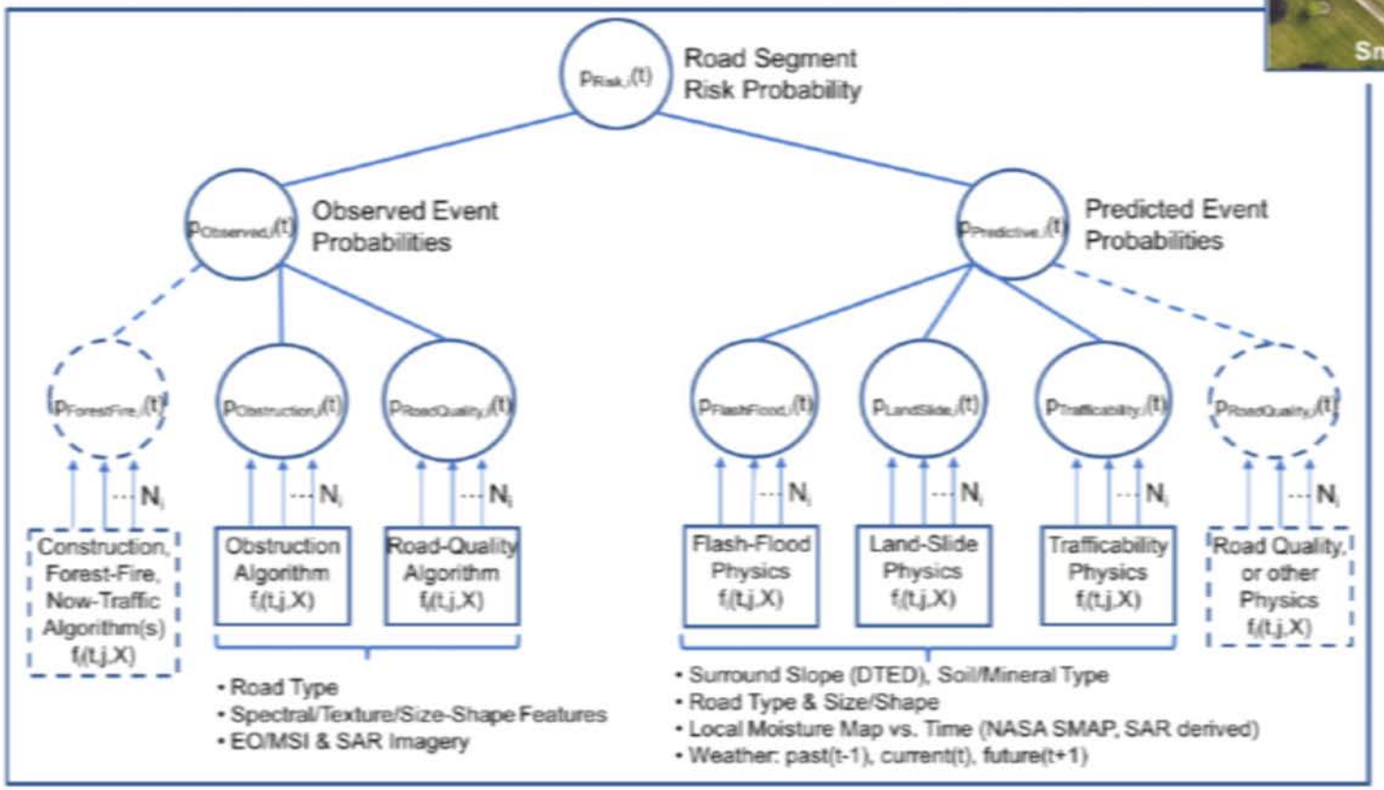


- In the general I&W OCEAN framework:
  - Objects are road impediments, scene changes, and hydrological events
  - DSM ensures only use of important, high confidence data sources

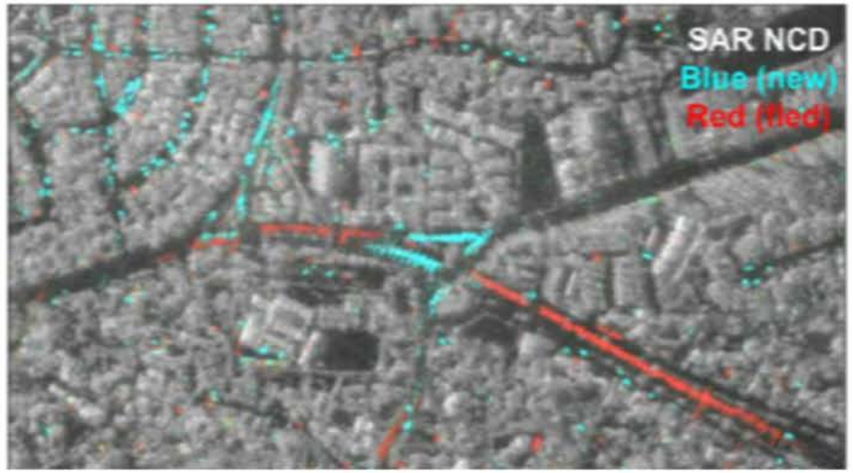
Trafficability: SRI approach

SURPAS: SUpply Route Predictive AnalyticS

- DBN space/time integrates both observed and predicted events to quantify risk level along routes of interest



- Road Quality: EO/MSI/SAR extracted spectral, texture, & shape features
- Obstructions: EO/MSI/SAR based changed detection & localization



Global digital elevation model: VSI approach

TA-2 Open Application

Global Disaster Assessment from Digital Elevation Models (DEMs)

- VSI DEMs from satellite imagery
- Compare DEMs from different epochs
- Change boundaries yield quantitative disaster assessment
- Meaningful descriptions (hospital, school, airfield) from geospatial feature databases

Satellite Imagery

DigitalGlobe AIRBUS DEFENCE & SPACE planet.

Open Source

OpenStreetMap tomnod GeoNames GlobeLand30 SpaceNet Semantic Web

ID	Location	Volume	Status	Description
1	(Lat,Lon)-1	1200 m^3	Demolition	Warehouse
---	---	---	---	---
---	---	---	---	---

2018.09.18

Vision Systems Inc. • GCA Phase 1 • Kickoff

Summary

- Trends in geospatial data and computing
  - Big data getting bigger
  - Cloud-based repositories are feasible
- Geospatial Cloud Analytics
- Scalable Data Platform
  - Let experts focus on analyzing
- Analytics
  - Demonstrate ability for analytics to scale globally

Can we provide analysis services rather than pixels for DoD warfighters?

www.darpa.mil